DOCUMENT RESUME

ED 066 102 HE 002 952

AUTHOR	Charvat, Josef; And Others
TITLE	A Review of the Nature and Uses of Examinations in Medical Education.
INSTITUTION	World Health Organization, Geneva (Switzerland).
REPORT NO	WHO-PHP-36
PUB DATE	68
NOTE	80p.
AVAILABLE FROM	The American Public Health Association, Inc., 1940 Broadway, New York, New York 10019 ($1.75)
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Academic Achievement; *Evaluation Methods; *Higher Education; *Medical Education; *Medical Students; Physicians; *Student Evaluation

ABSTRACT

Those responsible for the health services of a country are concerned above all with the quantity and quality of the young physicians who graduate from the medical schools. Examinations of medical students provide medical teachers with feedback as to the quality of their students. This document presents a review of present examination practice in different areas, methods of examination in current use, and new developments in examination theory and practice. This last section includes discussions of: (1) the process approach to determining what a test measures; (2) critical requirements approach to determining what should be measured; (3) new techniques for determining the full range of professional competence; (4) new approaches to the reporting and analysis of examination data; (5) new approaches to the problem of setting standards of competence; and (6) new developments in the training of examiners. (HS)

PUBLIC HEALTH PAPERS **36**

A REVIEW OF THE NATURE AND USES OF EXAMINATIONS IN MEDICAL EDUCATION

J. CHARVAT, C. McGUIRE & V. PARSONS

WORLD HEALTH ORGANIZATION
GENEVA

PUBLIC Health Papers *is a medium for the publication of occasional papers that have usually been prepared as contributions to the study by the World Health Organization of a particular health question, and that have been considered to be of interest to a wider circle of readers than those for whom they were originally written.*

The purpose of Public Health Papers *is to stimulate international thinking, discussion, and planning by the publication of the personal ideas, observations, and suggestions of individuals or groups.*

Reports of work completed under the auspices of the World Health Organization and recommendations of formally constituted international groups are to be found in the Organization's other publications.

A French edition of Public Health Papers *is published under the title* Cahiers de Santé publique. *Editions are also published in Spanish under the title* Cuadernos de Salud Pública, *and in Russian under the title* Tetradi obščestvennogo zdravoohranenija.

PUBLIC HEALTH PAPERS

No. 36

A REVIEW OF THE NATURE AND USES OF EXAMINATIONS IN MEDICAL EDUCATION

A REVIEW OF THE NATURE AND USES OF EXAMINATIONS IN MEDICAL EDUCATION

JOSEF CHARVAT

Chairman, and Professor of Medicine,
Charles University, Prague, Czechoslovakia

CHRISTINE McGUIRE

Chief, Evaluation Studies Section,
Center for the Study of Medical Education, University of Illinois,
College of Medicine, Chicago, USA

VICTOR PARSONS

Senior Lecturer in Medicine,
King's College Hospital Medical School,
University of London, England

WORLD HEALTH ORGANIZATION

GENEVA

1968

4

PRINTED IN FRANCE

CONTENTS

PREFACE

Those responsible for the health services of a country are concerned above all with the quantity and quality of the young physicians who graduate from the medical schools. One of the most effective methods of measuring quality is the evaluation of the students' academic performance by means of examination techniques. The older examination systems suffered from the disadvantages that they were not sufficiently objective and often conditioned the student to memorize only those facts that he believed would best satisfy the examiner. There has therefore been a search for new techniques of evaluation based on scientific principles.

During the last twenty years, impressive advances have been made in the behavioural sciences, educational psychology and statistics. Research carried out by educationalists both inside and outside medical schools has made it possible to define a number of fundamental requirements without which evaluation and grading systems cannot be considered valid, reliable and discriminating. As a result, medical educational concepts in general and techniques of evaluating student performance in particular are undergoing considerable changes.

Understandably, innovations in examination techniques have not always met with ready acceptance. There is, however, growing realization that these new techniques are devices to motivate and stimulate learning, to grade students more reliably, to provide a better insight into the didactic abilities of the teaching staff, and to shorten the time-consuming procedures of correcting essays and attending oral examinations. Finally, the newer evaluation methods permit reliable comparisons to be made of the academic attainments of students, not only within the class and the faculty but also in different medical schools and, as evidenced by studies already initiated in three continents, at the international level.

In order to obtain a comprehensive picture of the present situation, WHO decided in 1966 to undertake a review of both the old and the new procedures for the evaluation of student performance, with special attention to advances in examination techniques. Professor J. Charvat (Czecho-

slovakia), Miss C. McGuire (USA) and Dr V. Parsons (United Kingdom) co-operated in this study and their report is presented in the following pages. It is hoped that it will help teachers in medical schools to become familiar with these recent developments and to appreciate the value, limitations, and potentialities of the different examination techniques.

CHAPTER 1

INTRODUCTION

Examinations in medical schools may be viewed from different aspects. From the community point of view they represent a means by which society seeks to ensure a sufficiently high standard among those who will be permitted to act as physicians.

From the student's point of view they are barriers to becoming a doctor which have to be overcome at all costs.

From the teacher's point of view, the function of examinations is sometimes confused. In some schools, they are used as a check on the student's work, or they may be used as a device to reduce the number of students in overcrowded classes. Others would rather regard the examination as a device to give information about the student's learning. The examiner's role, then, can be looked at as changed from that of a judge to one of a counsellor, using the information gained from the examination as a basis for specific advice to the student.

The results of examinations thus also provide "feedback" information to the faculty. This is useful in assessing the efficiency of the teaching staff's efforts and methods, and in helping the student to assimilate the information that has been presented to him in a variety of ways.

Although most medical schools in the world feel the need to revise their curricula, not many of them have initiated systematic and scientific research into the nature and evaluation of their teaching and examination methods. A better understanding of the complex relations between the teacher and the student and between the teaching-learning process and examinations is still needed.

In all countries, the leading medical schools bear the mark of local culture, tradition, and needs. Considerable diversity both in teaching and examinations is, therefore, to be expected between one country and another. On the other hand, the increasing international mobility of populations and physicians makes it desirable to find out if some standard tests can be applied to students all over the world for purposes of comparing the results of various types of medical education.

The authors have tried to identify some of the principles that could eventually be accepted if properly championed.

The present report deals with the type of general medicine that has evolved in Europe and North America. No effort has been made to include such specialized medical schools as, for example, the paediatric faculties and the faculties of public hygiene existing in some socialist countries, or to consider the ancient traditional medicine practised in some regions of Asia and elsewhere.

CHAPTER 2

A REVIEW OF PRESENT EXAMINATION PRACTICE IN DIFFERENT AREAS

SELECTION OF STUDENTS

In all countries, minimum standards of age and length of secondary education must be met before a student can be considered for entry into a medical school. In some countries, the requirements also include specific standards of education in the humanities and science. The required standards are usually set by the examination boards of the schools, and sometimes by colleges or universities. The school leaving certificate (general certificate of education, maturity certificate, *Abitur*, matriculation certificate, *baccalauréat*, or *attestat zrjelosti*) has some reference to the standards imposed by universities for entrance.

In some countries, this certificate is sufficient to gain admission to a premedical year in which further selection takes place on the basis of written and/or oral tests. Many countries have replaced this costly selection procedure by devising admission tests for the selection of the most suitable students from the large number of applicants. Such admission procedures are sometimes based only on the grades obtained in the matriculation examination, supplemented by an interview and/or headmaster's report. However, many medical schools now insist on written tests in basic science subjects followed by an oral examination before the final interview. In the USA, most medical colleges require applicants to take the Medical College Admissions Test (MCAT), but this does not necessarily imply that they all have the same minimum standard of acceptance. Whatever the nature of the admission tests, performance in these is generally considered together with evidence from interviews and records of prior scholastic attainment.

It is becoming increasingly difficult for a student who has studied the arts and humanities only and who therefore has no basic scientific qualifications to gain entrance to a medical school. In an effort to alleviate this problem, a basic science course is provided by the university or the medical school in some countries so that the necessary scientific qualifications mays be acquired after provisional acceptance.

EVALUATION OF THE STUDENT'S PROGRESS THROUGH PREMEDICAL AND PRECLINICAL YEARS

Examinations at these stages are divided into two types:

Type A.—Course certificates, examinations "for the record", departmental examinations

These examinations are usually arranged with varying frequency to test a student's progress as each area of the syllabus is covered. They may be conducted for a group of students as, for instance, in departments of anatomy where, as study of each area of the body is completed, the student is "signed up" as having achieved a satisfactory standard. Such examinations may be given weekly, at the end of a course, or at the end of each term or semester. In other subject areas, oral examinations may be supplemented by written and practical examinations. In a minority of schools in Europe, multiple-choice examinations are administered.

The responsibility for the frequency and type of examination usually rests with the department concerned. Although external examiners are rarely engaged, interdepartmental and interdisciplinary examinations are emerging as a means of reducing the multiplicity of examinations a student may be required to take during the two or three years of his preclinical studies.

Satisfactory results in these departmental examinations may be all that is necessary for promotion into the clinical years in many schools in North America; in other parts of the world they are regarded as evidence of satisfactory attendance and performance and make the student eligible to progress to the major examinations held at yearly or two-yearly intervals.

Type B.—Major intermediate examinations (for example, the second M.B., Physikum, gosudarstvennyje)

Major examinations are not used extensively in many areas of the world, reliance being placed more upon the frequent tests described above. Under these circumstances, importance is given to marks reported by different departments in order to obtain an overall picture of the student's ability. If major examinations are used they can become a bar to advancement into the clinical years, and failure at this stage may delay the student by as much as six months or a year; repeated failure may lead to his being excluded from the medical course.

As currently employed, the major examinations typically include an essay, practical and oral components and, in some countries, multiple-choice examinations as well.

Excellence in these examinations may be rewarded by offering to the medical student a position as an *interne des hôpitaux* (France) or it may be one of the factors entitling him to an extra year's study in a preclinical subject leading to a science degree (United Kingdom).

EXAMINATIONS IN THE CLINICAL YEARS LEADING TO GRADUATION AND LICENSING

The frequency and type of testing show great variety; in some schools, the evaluation of achievement is based on frequent observation and the questioning of students in the clinics and in the wards, while written examinations are kept to a minimum. External examiners are sometimes utilized and may include professors from preclinical and clinical disciplines, or physicians in general medical practice.

In some medical schools in the United Kingdom and in the USA two years may pass before a student sits for any official examination. In most other areas, written and oral examinations are held at the end of each year or course of instruction. The majority of these examinations are arranged entirely by the faculty or department, and in some countries (Sweden, for example) the student can take them at his own pace and at a time agreed between the professor and himself. In other countries, the student is examined in four to twelve subjects, and he is required to take these examinations within a few weeks. These are normally written and oral examinations, and tests of practical performance at the bedside. However, in various areas a more objective, multiple-choice type of examination (M.C.E.) is gradually being introduced, and in the USA films and patient management problems[1] are beginning to replace written and oral types of examination.

In some countries, a graduation degree in medicine does not automatically give the physician the right to practise where he wishes, and compulsory prelicensing or preregistration years in hospital practice are necessary. When the physician has met this requirement, he may be licensed unconditionally or may be required to take a national or state licensing examination. In many areas, further years of study and the writing of a thesis, which may have to be defended in a conference open to the public, are required before the degree of M.D. is given. In only a few countries are presentation of, and examination on, a thesis a necessary supplementary qualification before the final examination.

[1] See Chapter 4, Development of appropriate tests in the cognitive and psychomotor domains, and Annex 3, p. 58.

CHAPTER 3

METHODS OF EXAMINATION IN CURRENT USE

An examination is a complicated psychological and social interaction between the examiner and the pupil. This interaction is influenced by many subconscious factors on the part of the student (e.g., emotional state) and by the temperament and character and, on occasion, the level of professional competence of the examiner. Undesirable factors influencing the examiner's biases should, of course, be minimized (or entirely eliminated) by:

(*a*) repeated self-analysis through available feedback mechanisms;

(*b*) increased objectivity through the substitution of programmed examinations based on multiple-choice and short written answers and/or increased standardization of some of the conventional examinations.

In the survey below we hope to show that a thorough appraisal of examination methods in common use is required in order to obtain an accurate estimate of the areas of professional competence that are now being assessed.

CRITERIA FOR COMPETENCE-MEASURING TECHNIQUES

Objectivity

Any techniques for measuring medical competence must yield objective data, i.e., independent observations of different experts must agree. (An examination is "objective" when, for example, different examiners independently arrive at closely similar grades for each of a series of essays or oral examinations; or when different experts independently select the same alternative as the best answer to each of the multiple-choice questions that comprise a test.) In general, objectivity

is a function of the clarity and explicitness of the criteria used in making an observation or judgement. Whenever different observers attend to different attributes of a performance or attach different weights to these several attributes, objectivity will be impaired.

Validity

The examination must also be valid; in other words, it must measure what it purports to measure. For example, if a test purports to measure the student's ability to solve problems, it cannot be regarded as valid if the student need only search his memory in order to perform satisfactorily.

Reliability

Examinations are considered to be reliable if they yield stable or consistent scores when given repeatedly to the same group under similar conditions. Sampling errors are one source of unreliability of an examination. However, there are other factors that can make an examination unreliable, such as differences in the conditions under which the examination is held or in the health status of the examinee. Reliability is usually expressed in terms of a "reliability coefficient", which indicates what proportion of the test variance is non-error variance, i.e., due to true individual differences as opposed to sampling errors.

Techniques have been developed for the analysis of errors in examination scores, such as the test-retest technique, which consists in administering the same examination to the same group on two separate occasions. Another procedure, called the alternate-form technique, consists in administering two examinations closely similar in content and in their demands on intellectual ability.

Further, there is the split-half reliability method in which the examination is divided into two sub-examinations of equal length. The scores obtained in the two sub-examinations are then correlated and, if necessary, corrected according to the Spearman-Brown formula.[1] Another technique uses the Kuder-Richardson formulae for computing the reliability of examinations.[2]

[1] Angoff, W. H. (1953) Test reliability and effective test length. *Psychometrika*, 18, 1-14.
[2] Hoyt, C. J. (1941) Test reliability obtained by analysis of variance. *Psychometrika*, 6, 153-160.

THE PRESENT USES OF EXAMINATIONS

The following types of examination are at present in use in many areas of the world:

1. *Examinations to select students for entry into medicine*

In order to avoid having too many failures and to keep to a minimum the waste of human resources it is important for the medical faculty to try to predict the applicant's chances of success.

Experience suggests that selection is best based on consideration of the following:

(*a*) an estimate of past academic performance based on the candidate's achievement in the premedical curriculum;

(*b*) an estimate of the motivation of students based on references, interviews and other information about the candidate's personal qualities;

(*c*) an estimate of the candidate's ability to pursue the medical programme, using reliable and valid tests of required aptitude and intelligence.

There is no agreement as to which is the most reliable and valid type of assessment. Even in the USA, although the Medical College Admission Test (MCAT) is widely used it still has many critics. The test consists of four parts designed to measure:

(*a*) ability to manipulate verbal symbols;

(*b*) ability to manipulate quantitative symbols;

(*c*) achievement in science; and

(*d*) achievement in social and behavioural sciences.

Experience with this test over the past few years has indicated that among students with relatively low scores the risk of failure in the medical course is considerably increased. There is growing awareness that tests of the MCAT type measure only a few of the important prerequisites for success in a medical school. For this reason, it is recommended that such a test be used in conjunction with other evidence in the selection of medical students.

Defects and abuses

(*a*) The too rigid application of aptitude tests without due regard for other important factors in student success (mainly motivation, work

habits and study skills, quality of academic preparation and the ability to think independently) may lead to biased selection.

(*b*) Given the current rate of change in medical science and the requirements for a variety of specialists in the health field, the results of aptitude tests must be used flexibly. This, however, may not be sufficient. Rather, the aptitude tests themselves have to be designed flexibly, i.e., with a stronger bias on medical care and preventive and community medicine, or with more emphasis on basic medical sciences in the case of schools that aim at superspecialization or research orientation of their students.

(*c*) The age of the applicants has to be taken into consideration, as, for instance, the level of maturity found in the 16-17 year old age group is different from that found in the 18–19 year old group.

2. *Examinations to assess student progress and to guide further learning*

Once the student has been admitted to the medical programme it seems essential to provide him and his instructors from time to time with systematic and accurate appraisals of his strengths and weaknesses in order to guide his further education most efficiently and effectively. Such examinations should be designed:

(*a*) to evaluate the student's progress;

(*b*) to motivate the student by techniques that involve both encouragement and reproof; and

(*c*) to select for excellence as well as for minimum competence and to assist in determining the student's aptitude or eligibility for extra degree courses, residency appointments in hospitals, or special awards.

Defects and abuses

(*a*) Many examinations are poorly prepared in that they test minutiae instead of principles or sample only a narrow range of the requisite knowledge and skills.

(*b*) Interim examinations "for the record" are sometimes used as a method of interdepartmental competition for the student's time and attention. The experience of instructors lecturing to a nearly empty hall because students have been distracted by imminent examination requirements of a competing course is almost universal.

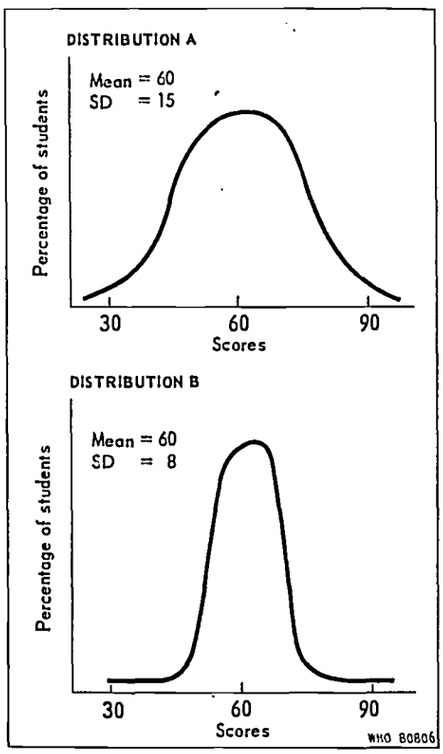
(*c*) Occasionally examinations may be too frequent. They then destroy any capacity for independent learning and thus encourage the cramming of knowledge to be recalled and soon to be forgotten.

(*d*) Interim examinations in some countries are used to screen out unsuitable candidates when there has been little or no initial selection.

(*e*) Some examinations are not designed to assess the full range of competence (from satisfactory to superior). This is the case if they contain only very difficult or only very simple questions.

A test is regarded as discriminating if there is a wide range of test scores. However, it is not enough for a few individuals to be at the extremes. For example, in the accompanying figure, discriminations throughout the range would be better on the test with distribution A than on that with distribution B.[1]

TWO TYPES OF TEST SCORE DISTRIBUTION



[1] In test analysis the test-maker is also concerned with how useful each item is in separating the good students from the poor ones. *An item* is regarded as discriminating if it is answered correctly by more of the good students (i.e., those with high scores on the test) than by poor students (i.e., those with low scores on the test).

3. *Examinations designed to permit certification of satisfactory standards of the student's knowledge*[1]

It seems obvious that it is necessary to design an appropriate variety of examinations (based on both theoretical and practical exercises) to determine whether or not the student meets appropriate standards of competence and, if so, in which subject he has distinguished himself.

Indeed, a detailed programme of student assessment must be one of the objectives of the faculty of the medical school, which has to decide on:

(*a*) the amount of knowledge the examinee must have;

(*b*) what types of problem he has to solve;

(*c*) what technical skills he should have developed;

(*d*) what professional attitudes and habits he should have acquired in general and toward the patient in particular;

(*e*) to what extent he should have demonstrated a capacity for original work; and

(*f*) to what extent he should have developed an ability to initiate small research projects.

In the light of the above criteria, to be applied when certifying the student's knowledge after his final year of study, it seems appropriate to look into the errors or omissions made when designing such examinations.

Defects and abuses

(*a*) It appears from a review of present practice that too much reliance is placed on the evaluation of a very limited aspect of the student's knowledge, primarily the ability to recall isolated fragments of information.

(*b*) The standards used in assessing a given aspect of the student's knowledge are often determined by a single individual or department in relation to specialist requirements and may be totally inappropriate for the general physician.

(*c*) As there are usually no accepted standards for any particular student group or examination period, the specific standards that a student is required to meet may be determined fortuitously by the time at which he takes the examination and by the examiner to whom he is assigned.

[1] For example, the examination leading to the award of a university degree, or the final year examination in medicine (State examination).

(*d*) In some countries examinations, combined with other evaluations of the student, are used to make distinctions between students that go far beyond the limits of confidence of the measurements made.

CURRICULAR ALLOCATIONS

Evidence derived from a study of the performance of students with respect to all essential criteria of the student's knowledge can provide a valuable basis for readjusting the curriculum, e.g., the teaching time allocated to different departments. If there is evidence that certain goals are not being reached, e.g., in public health problems, an effort can be made to remedy this deficiency by allotting more curricular time to this area and, for example, by changing the instructional methods or by initiating an interdisciplinary programme to include the study of these problems.

A detailed description of student performance with respect to all major aspects of competence would provide an institution with evidence regarding the effectiveness of its current programme and with sound data that could be used to assess the relative merits of alternative programmes or instructional methodologies. Such scientific evidence is essential as a guide to rational decision-making. However, such evidence is rarely gathered; when attempts are made to assemble it, they are not carried out systematically and are not associated with a well-planned overall evaluation programme. For the most part, evidence used in assessing the distribution of curricular time and proposed curricular changes is derived from general impressions rather than from a detailed analysis of student performance. Generally, assessments are based on a rather narrow definition of the student's knowledge rather than on the set of abilities required for the contemporary physician to perform successfully his varied professional tasks.

RECOMMENDATIONS

We recommend, first, that systematic consideration be given to a re-definition of all the essential criteria, on how and to what extent to measure the student's knowledge. As mentioned before, these should measure skill in solving problems, ability to communicate with patients, colleagues and other members of the health team, and other professional requirements.

Second, we recommend that the standards that the student is expected to meet be developed by the faculty acting in concert and not by any

uncoordinated decisions of medical specialists acting independently. The views of the students could, at that stage, be of considerable help to the faculty.

Third, we recommend that attention be given to developing examinations that measure the full range of the student's knowledge and that the student's performance be assessed according to the standards set by those who were responsible for the design of the examination.

ADVANTAGES AND DISADVANTAGES OF DIFFERENT TYPES OF EXAMINATION

Present methods of examination can be divided into six main categories:

(*a*) oral examinations
(*b*) practical examinations
(*c*) essay examinations
(*d*) objective examinations:
  (i) multiple-choice, constructive or selective-type questions
  (ii) completion-type questions
(*e*) observational reports on student's performance
(*f*) theses and research projects.

The major advantages and disadvantages of each of these methods are set out below:

(*a*) *Oral examinations*

Purpose: To permit the student, through his answers to questions put to him orally, to demonstrate his knowledge and understanding in his subject of study, as well as his thinking and problem-solving ability.

Disadvantages	*Advantages*
1. Inadequate standardization.	1. Direct personal contact with candidates.
2. Insufficient objectivity and reproducibility of results.	2. Opportunity to take into account mitigating circumstances.
3. Possible abuse of personal contact with examiner and probably cueing.	3. Flexibility in moving from strong to weak areas.
4. Undue influence of irrelevant factors.	4. Opportunity to ask the candidate how he arrived at an answer.
5. Few trained examiners available.	5. Opportunity for simultaneous assessment by two examiners.
6. Excessive cost in professional time in relation to the limited value of the information obtained.	

(*b*) *Practical examinations*

Purpose: To reveal what the examinee *can* do as distinct from what he *says* he can do.

Disadvantages

1. Insufficiently standardized conditions, whether in laboratory experiments using animals or in bedside examinations with patients of varying degrees of co-operativeness.
2. Insufficient objectivity and intrusion of irrelevant factors.
3. Limited feasibility for large groups.
4. Difficulties in arranging for examiners to observe candidates demonstrating the skills to be tested.

Advantages

1. Opportunity to test skills involving all the senses with observation of performance by examiner.
2. Opportunity to confront the candidate with new problems, both in the laboratory and at the bedside, to test his investigative ability as distinguished from his ability to carry out "cookbook" exercises.
3. Opportunity to observe and test attitudes and responsiveness to the total situation.
4. Opportunity to test the ability of the student to communicate with the patient, to discriminate between important and trivial issues, to arrange and display the data.

(*c*) *Essay examinations*

Purpose: To permit the examinee to give in writing and in his own words a relatively free and extended response to a problematic situation and thus to reveal information regarding the student's mental processes.[1]

Disadvantages

1. Severe limitation of the area of the student's achievement that can be sampled.
2. Difficulties in obtaining objective judgements of performance.
3. Negligible feedback to the student.
4. Excessive time required for scoring.

Advantages

1. Opportunity to test not only a candidate's store of information but also his ability to organize ideas and express them effectively in his own language.

(*d*) *Objective examinations*

Purpose: To permit different examiners independently to arrive at the same or very similar grades for each examination question. These examinations are of two types: (i) The multiple-choice question, consisting of an item stem, either in the form of a direct question or an

[1] Harris, C. W., ed. (1960) *Encyclopaedia of educational research*, 3rd ed., New York, Macmillan.

incomplete sentence, and a number of responses, one of which is the best answer; the other answers are referred to as distractors. (ii) The completion type question, where one or several key words in a given sentence have to be filled in.

Disadvantages

1. Construction is time-consuming if arbitrary and ambiguous questions are to be avoided.
2. Necessity of making allowance for positive scores that may be achieved by guessing.
3. Much prejudice among teachers against this type of examination.
4. Cues are provided that are unavailable in practice.

Advantages

1. Significant increase in the range and variety of facts that can be sampled in a given time.
2. Opportunity to test the candidate at the desired level by varying the difficulty of the questions and, in the case of multiple-choice questions, including as possible answers misconceptions common at his level of training.
3. Opportunity to obtain detailed feedback for both student and faculty.
4. Very economical for large groups.
5. The standards of scoring can be kept constant for many years.

(*e*) *Observational reports on student's performance*

Purpose: This type of test serves the dual purpose of identifying those with exceptional abilities and/or revealing those with persistent intractable deficiencies with respect to professional conduct and attitudes.

Disadvantages

1. The examiner acts as both observer and judge.
2. Extended contact with student required for a valid estimate of his performance.

Advantage

Opportunity to obtain fuller and usually more valid information about a candidate; by pooling the reports of many examiners the results can be made more reliable.

(*f*) *Theses and research projects*

Purpose: This type of assignment is designed primarily to provide information on the student's ability to collect information and to put it in order; he is normally expected to work independently and at his own convenience. However, we believe that it is not appropriate for all students and should be treated as an optional type of examination. Therefore, we feel that a list of advantages and disadvantages can be omitted.

GRADING SYSTEMS

In the more objective type of test, grading is no longer a problem provided that agreement can be reached on what constitutes the correct answer and on the formula to be applied in weighting correct and incorrect answers and in making adjustments for guessing. However, in the essay-type examination, safeguards must be introduced in order to make the results meaningful for the student and the department involved. This can be achieved by:

(*a*) having examiners determine in advance the essential features that must be included in a satisfactory answer and the importance to be attached to the style, organization, logic and synthesis of the material included in the answer;

(*b*) arranging for independent grading and exchange of papers between a minimum number of examiners. We disapprove of allowing one examiner alone to determine the grade on the paper as a whole or on any section of a final examination.

A variety of marking or grading systems are in use throughout the world, but, whatever the system of number or letter scores employed, it is ultimately necessary to decide whether the candidate has failed an examination or not. It is also necessary to decide whether or not and, if so, how, he can compensate for a deficiency. For such decisions, "rules-of-the-game" must be established to deal with the isolated "crucial" answer that fails the student and with the evasions of candidates who misconstrue questions. The major deficiency in a grading system with only the pass-fail difference is that it yields insufficient information about the candidate's specific strengths and weaknesses. Thus, it provides little feedback to interested departments, schools or counsellors as to why a particular candidate or programme failed to reach a predetermined goal. Consequently, there is little or no information about what should be done in the future, either in guiding that candidate's education or in improving the programme for a majority of the students.

CHAPTER 4

NEW DEVELOPMENTS IN EXAMINATION THEORY AND PRACTICE

> While learning is the objective of teaching, and while the teacher is a major instrument for its facilitation, evaluation provides the final evidence of whether learning has been accomplished and some insight into whether the teacher was effective. Although the advances in the field of evaluation during the last 25 years have been both substantial and significant, the tools of evaluation that are most widely used in most parts of the world were already old a century ago. Medical teachers can no longer fulfil their educational responsibilities adequately without more knowledge than most now have of the criteria by which they can select, from the increasingly varied array of evaluation tools, those that will provide the most valid and reliable data on the kind of behaviour they are attempting to assess.[1]

The new developments in testing designed to provide more valid and reliable data are described below with regard to:

A. Methods of determining what a test measures
B. Critical requirements for determining what should be measured
C. Techniques for measuring the full range of medical competence
D. Methods of reporting and analysing examination data
E. The problem of setting standards of competence
F. The training of medical teachers

A. THE "PROCESS APPROACH" TO DETERMINING WHAT A TEST MEASURES[2]

In the "process approach" to analysis of what an examination measures the examination (oral or written) is described in terms of the

[1] WHO Expert Committee on Professional and Technical Education of Medical and Auxiliary Personnel (1966) *Fifteenth report: The training and preparation of teachers for medical schools with special regard to the needs of developing countries* (*Wld Hlth Org. techn. Rep. Ser.*, No. 337).

[2] For a description of some of the commonest weaknesses of current examinations, together with some suggestion for remedying them, see Annex 3, p. 55.

intellectual (or other) abilities that are required to respond to each of the questions or problems posed. A system of classification representing intellectual (or other) processes ranging from "simple recall of isolated information" to "complex problem solving" (see "Critical requirements approach" below) is used to categorize *each* question. Experts in the subject matter consider each question separately and attempt to determine by introspection what intellectual process an individual (at the level of education and experience for which the test is designed) would need to use in order to answer the question. The examinee may simply have to search his memory. If he can "figure out" the answer, how does he go about it? More than simple recall may be required, and the examinee may have to show that he recognizes the meaning of a fact or concept. He may be required to formulate (or select) a relevant generalization to explain a particular phenomenon. He may be required to interpret data, to apply general principles, to evaluate a total situation, or to make a decision about a complex problem.

The various studies done to date in which this approach has been used[1] reveal that the overwhelming proportion of questions (75-95%) in the examinations currently in use in the USA and Canada measure only the *recall* of information. The form of the examination did not influence this finding, as "recall of information" was equally characteristic of the oral, essay and objective-type examinations studied.

A second approach to the analysis of examinations has been an attempt at empirical verification of the "process approach", either through interviews with students to determine the intellectual process they do in fact employ in answering specific questions,[2] or by means of correlational and other statistical studies of the attributes of examinations purporting to measure different types of intellectual competence. In one such series of studies, conducted by the Office of Research in Medical Education at the Center for Study of Medical Education, University of Illinois College of Medicine (unpublished reports, 1963 through 1966), it was found that correlations between scores on sets of questions carefully designed to measure interpretation of data or clinical problem-solving and scores on tests of recall rarely exceed 0.40 and more commonly vary between 0.20 and 0.33.

Finally, without regard to the "process approach" *per se*, various statistical techniques, including factor analysis, have been used to analyse the number and types of intellectual (or other) factors sampled

---

[1] McGuire, C. (1963) *J. med. Educ.*, 38, 556. See also Annex 4: Canada, p. 61.
[2] Bloom, B. S. & Broder, L. J. (1950) *Problem-solving processes of college students; an exploratory investigation*, Chicago, University of Chicago Press.

in the multiple measures employed in student assessment. In one of the most definitive of such studies Schumacher[1], analysing the results from 306 medical students in five medical schools in the USA, concludes:

> The majority of measures used to assess medical student accomplishment measure a single, general dimension which might be labelled "General Medical Knowledge" (Factor 1). This appears to be a complex dimension that is reflected in faculty judgments, extra-mural examinations and judgments made by fellow students. In addition to "General Medical Knowledge", a set of personal characteristics that might be called "Skill in Patient Relationships" (Factor 2) can be measured by certain peer ratings and, *to some extent*, by fourth year grades in medical school.

B. CRITICAL REQUIREMENTS APPROACH TO DETERMINING WHAT SHOULD BE MEASURED

Clearly, knowledge of a vast quantity of information is a prerequisite for satisfactory performance as a physician but, in itself, is not sufficient to assure competence. Careful studies have repeatedly demonstrated very low correlations (often not significantly different from zero) between scores on tests that measure the ability to recall information and tests that measure other intellectual abilities or professional skills. The development of a rational programme of student evaluation and the selection of appropriate examination techniques to implement that programme therefore require that a series of decisions be made to indicate precisely the total range of qualities that should be assessed. Application of this principle to medical education implies that it is necessary, as a first step, to define the professional responsibilities of the physician in the light of the health needs and the organization of health services in his geographic area. Consideration (including systematic empirical analyses) of the qualities that make for outstanding performance (rather than those that make for unsatisfactory performance) in discharging medical responsibilities will show that the requirements for medical competence fall into three main categories:

(1) those in the cognitive domain (e.g., knowledge, understanding, problem-solving ability)

(2) those in the psychomotor domain (e.g., technical skills)

(3) those in the affective domain (professional attitudes, habits, values).

---

[1] Schumacher, C. F. (1964) *J. med. Educ.*, 39, 192.

It is clear that in each of these domains there are certain critical requirements that affect physician performance. It follows that in any fully effective programme of student evaluation each of these critical requirements should be specified, and a method developed for measuring the extent to which the student has to meet them. Chart 1 below illustrates such a table of specifications.

CHART 1

AN ILLUSTRATIVE LIST OF CRITICAL PERFORMANCE REQUIREMENTS FOR PHYSICIANS [1]

I. Cognitive domain

1. Knowledge of fundamental vocabulary, facts, concepts, principles, laws, methods and procedures
2. Understanding of these facts, concepts, etc.
3. Ability to understand and interpret data
4. Ability to solve relevant problems
5. Judgement in evaluating a total situation
6. Ability to create a new synthesis

II. Psychomotor domain (technical skills, etc.)

1. Skill in questioning the patient in order to take a case history
2. Skill in performing physical examinations
3. Skill in using various laboratory and clinical instruments
4. Skill in making accurate observations

III. Affective domain (attitudes, habits, values)

1. Acceptance of responsibility for patient welfare
2. Concern and consideration for patient and patient's family
3. Recognition of medical capabilities and limitations
4. Ability to establish effective relationships with colleagues and other members of the health team
5. Regular observation of appropriate safeguards
6. An inquiring mind
7. Willingness to use medical capabilities to contribute to community as well as individual patient welfare.

[1] For additional definition of these requirements see Annex 2, p. 51.

C. NEW TECHNIQUES FOR DETERMINING THE FULL RANGE OF PROFESSIONAL COMPETENCE

1. *Development of specifications for an examination*[1]

As with the analysis of examinations, the "process approach" has been found helpful in the *construction* of more reliable and valid examinations of all types: multiple-choice, essay, oral and practical. The first step in applying the process approach to the construction of medical examinations is to determine in precise detail the specific type or types of intellectual (or other) competence the test is designed to measure. In determining what is to be measured, it is necessary to decide upon (*a*) the body of information and (*b*) the abilities and skills to be tested, i.e., (*a*) the range of facts, concepts, principles, and techniques the candidate should "know", and (*b*) what he should be able to do with this content (repeat it, interpret it, apply it to new problems, or seek extensions of it in the developing literature).

Once the content and the intellectual (or other) skills have been determined, it is helpful to define the particular behaviour that distinguishes the individual who has acquired them from one who has not. Specifically, what does the candidate who "applies" a principle to a new problem do that distinguishes him from one who cannot? For example, the candidate who is effective in interpreting data is able: (*a*) to read data presented in a variety of forms; (*b*) to translate data from one form to another; (*c*) to interpolate and extrapolate within the limits of the data; (*d*) to perceive significant relations among data; and (*e*) to determine the implications of the data. He is able to avoid: (*a*) crude errors of reading or interpretation; (*b*) going beyond the limits of the data; and (*c*) overcaution in interpreting the data.

2. *Development of appropriate tests in the cognitive and psychomotor domains*

(*a*) *New designs in conventional test formats*

The process approach to the evaluation of competence permits the widest possible latitude in devising "test situations". These may appropriately range from multiple-choice tests and questionnaires to objectively rated diagnostic or therapeutic interviews with an assigned

[1] For further discussion of this question, see Annex 3.

patient. Example 1[1] below illustrates a type of question that superficially and formally departs only slightly from conventional techniques. It requires the student to make a judgement of the type he will have to make as a physician, about clinical data presented in a realistic form: *Certain findings are made. What lesion could account for them and what are their practical consequences?* The student records his decisions in a fashion that permits both objective and reliable assessment of their accuracy. It is obvious that this method of formulating the question and recording the student's answer makes it very easy to identify the specific strengths and weaknesses of individuals and groups and to document improvement that has occurred over a time or as a consequence of curricular change.

EXAMPLE 1

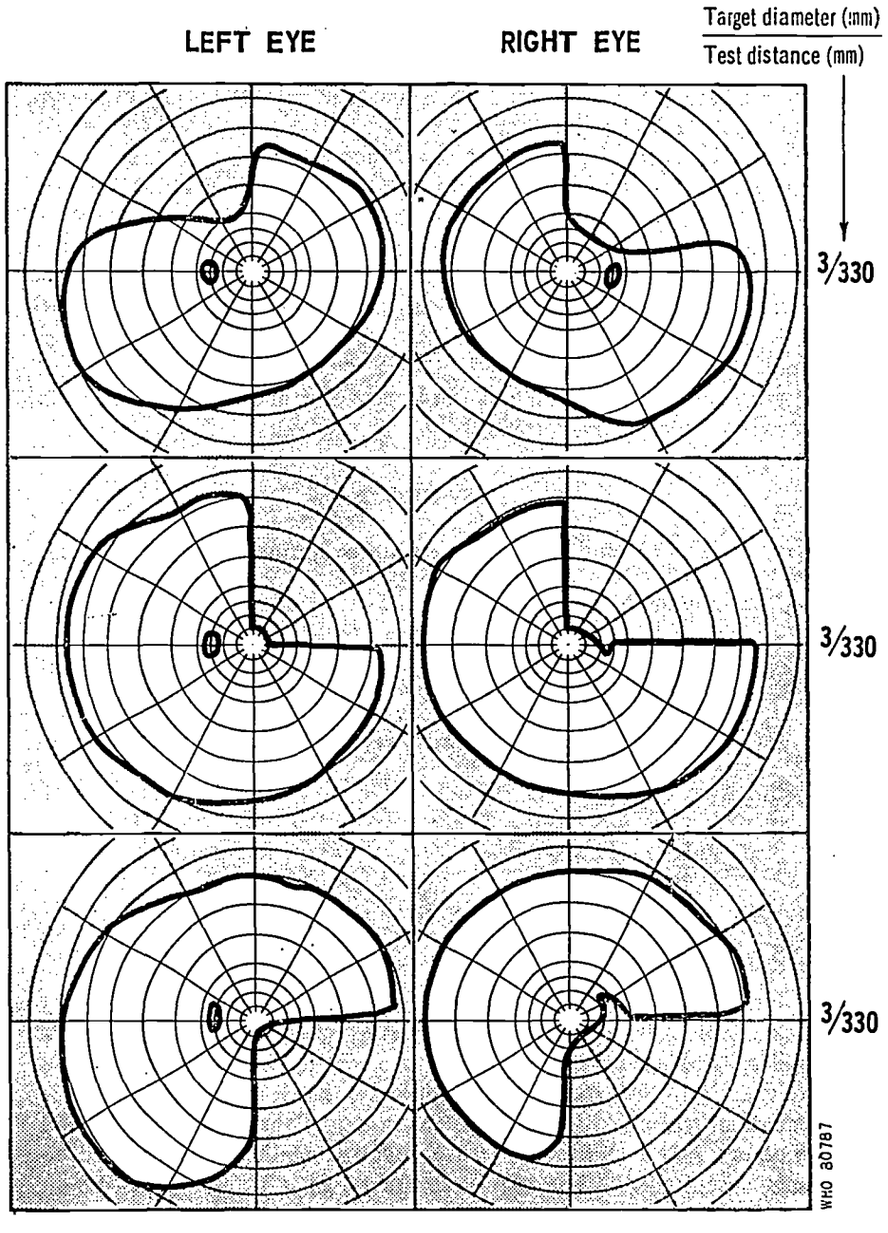
INSTRUCTIONS: In the appropriate space below name the most likely site of each of the visual field defects shown in the figure opposite. Also estimate (in terms of 0, + or ++) the handicapping effects of these visual field defects in each of two activities: (*a*) moving about as a pedestrian in heavy city traffic, and (*b*) reading.

Field defect No.	Most likely site	Handicapping effect	
		(*a*)	(*b*)
1			
2			
3			

Finally, in considering Example 1 it is important to observe that there are three possible alternatives: (*a*) to make the test objective by providing a series of possible alternative answers to each question from which the student is required to select the best one; (*b*) to allow the student to write in the answer, which he must formulate for himself, as in

[1] With the exception of Example 2, all illustrations are taken from the Comprehensive Examinations prepared by the Committee on Student Appraisal of the University of Illinois College of Medicine.

the example; or (*c*) to present the same questions orally to the student, who is provided with the diagrams. From the point of view of what the test measures, one technique has no special advantage over the other.



However, the completely objective test is by far the most economical of faculty time. The usual criticism to it is that it gives the student too much assistance by allowing him to select, rather than requiring him to formulate, his own answer. Studies indicate that this produces no significant difference in results when the completely objective form contains a sufficient number of carefully stated "wrong" alternatives (at least three and preferably more) that represent common misconceptions among students. Indeed the completely objective test has one distinct advantage: it permits the examiner to set the exact task and its intended level of difficulty with far greater precision than does any other type of test since, by careful formulation of the alternative answers, the examiner can control the exact degree of learning and discrimination he wishes the student to demonstrate. For instance, Example 2 shows two different ways of asking about a student's knowledge of the magnitude of the population of the USA. Example 2A requires far less precise knowledge than Example 2B and the examiner, by his choice of alternatives, effectively controls the level of discrimination that a correct answer to the question requires.

EXAMPLE 2

A. Which of the following is the best approximation of the current population of the United States?

(Circle the number of your choice)

1. 2 000
2. 200 000
3. 2 000 000
4. 20 000 000
5. 200 000 000

B. Which of the following is the best approximation of the current population of the United States?

(Circle the number of your choice)

1. 174 000 000
2. 176 000 000
3. 178 000 000
4. 180 000 000
5. 182 000 000
6. 184 000 000
7. 186 000 000
8. 188 000 000
9. 190 000 000
10. 192 000 000
11. 194 000 000
12. 196 000 000
13. 198 000 000
14. 200 000 000

Example 3 illustrates another method of questioning that requires the student to demonstrate that he can interpret clinical data presented in realistic form and that he can anticipate possible associated factors (causal or other). However, in contrast to Example 1, this method employs a new though rather costly modality, namely, the data obtained from auscultation. The test is conducted in a room equipped with

individual stethophones. High-fidelity tape recordings of heart, lung or abdominal sounds are reproduced through these stethophones in a manner that simulates as closely as possible the form in which the student is accustomed to hear them through his own stethoscope. He is required to identify what he hears and to indicate what might have been responsible for these findings or what condition is likely to be associated with them. Two elements in such tests are worthy of special note: (1) the student must make his decision on the basis of data presented in a form that closely simulates reality, *not* on the basis of a generalized verbal description; and (2) since the examiner can be sure that all students hear precisely the same sound, it is clear that this simulation of reality provides a degree of standardization that the real situation would lack.

EXAMPLE 3

Directions: You will now hear a series of heart sounds (A) and breath sounds)B). For each question circle the number of the *one best* answer.

Heart sound A is heard at the 2nd interspace to the right of the sternum, in a 50-year-old female.

(1) The basic cardiac rhythm is
1. Normal sinus rhythm
2. Sinus tachycardia
3. Sinus bradycardia
4. Extrasystoles
5. Auricular fibrillation
6. Bigeminy

(2) You can hear
1. A systolic murmur
2. A diastolic murmur
3. *Both* systolic and diastolic murmurs
4. *Neither* systolic nor diastolic murmurs

(3) Which of the following might produce these findings?
1. Hexamethonium
2. Digitalis
3. Nitroglycerin
4. Quinidine
5. Meprobamate
6. None of the above could produce these findings

Breath sound B is heard over the left lobus inferior dorsalis of the lung in a 34-year-old male.

(4) The breath sounds are
1. Bronchovesicular
2. Bronchial
3. Tubular
4. Amphoric
5. Not accurately described by any of the above

(5) It is possible to detect
1. Cracking rales
2. Bubbling rales
3. Musical rales
4. No rales
5. Friction rub

(6) These findings are likely to be associated with the condition found in
1. Slide 1
2. Slide 2
3. Slide 3
4. Slide 4
5. None of the foregoing

An increasing variety of laboratory, clinical and research data are now being utilized as a basis for questions in newer types of examinations.

For example, the student may be given a booklet of high-quality photographic reproductions including, for example, X-ray plates, pictures of patients, gross and histological specimens, optic fundi, photomicrographs of blood smears, and cultures. He may then be asked what findings are demonstrated in each photograph, what related findings he would expect, what might have produced the particular findings demonstrated, and so on.

For a test of the student's alertness to "findings that don't quite fit the clinical picture" as well as of his ability to interpret data presented in realistic form, a technique requiring interpretation of several types of data, as illustrated in Example 4, is employed.

EXAMPLE 4

Data about patient X: [A brief paragraph states the pulse, blood pressure, physical and auscultatory findings in an apparently healthy 16-year-old girl sent for examination because of a heart murmur discovered on routine physical check-up.]

Questions on patient X:

1. An X-ray plate of the chest is obtained. Which of the following is most consistent with the clinical history?
[Six X-ray plates are presented, and the student must select the most consistent one.]

2. Cardiac catheterization is performed. Which of the following would be the expected result?
[Six sets of specific results are described, and the student must choose the best one.]

3. After studies have been completed and a diagnosis has been established, therapy is recommended. Which of the following would be the most appropriate management?
[Six therapeutic plans are described, and the student must select the best one.]

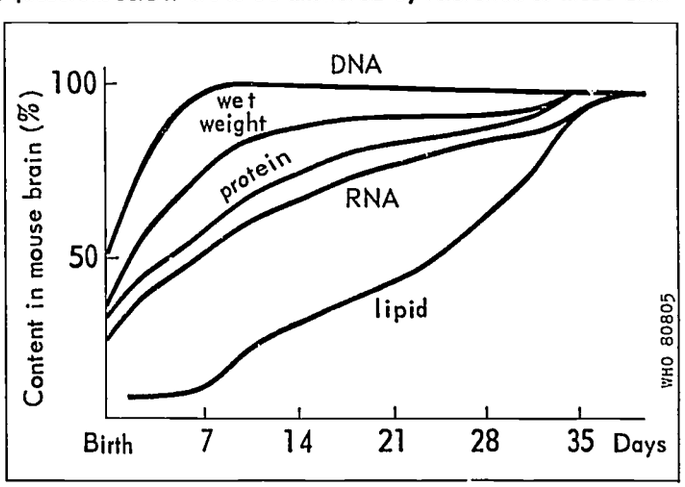
4. The prognosis for such a patient, if properly managed, is...
[Six alternative completions for this sentence are given, and the student must choose the one that best fits the specific case described.]

This technique of requiring the student to show that he can accurately interpret certain data may be extended to many other types of information. For example, brief colour films may be presented showing a patient walking across a room and in different positions, as well as close-ups of various aspects of the physical examination, after which the student is asked a series of multiple-choice or other objective questions designed to test his skill and accuracy of observation, his ability to anticipate related findings, and his judgement in planning the next appropriate steps in the diagnostic work-up of that specific patient. Since a substantial part of a physician's practice may be in the field of prevention,

it is of special relevance to make use of colour films showing the examination of well babies and children of different ages to determine whether the student recognizes normal limits for various age-groups or attributes pathology to every case. Sound films of excerpts from a psychiatric interview may be similarly used and questions asked about the patient's behaviour and the inferences that can be plausibly be drawn from it. Further, a good colour film of selected portions of a complete autopsy may be presented, together with other clinical data about the case, followed by a series of objective questions testing the student's skills of observation and interpretation and his understanding of the basic pathophysiologic processes demonstrated in the case. Electrocardiographic or electroencephalographic tracings or charts, diagrams, and figures containing experimental or epidemiological data may be printed in the test booklets, together with a series of questions requiring varying levels of discriminatory ability for their interpretation. Example 5 illustrates one such use of laboratory data.

EXAMPLE 5

The graph below shows the increase in wet weight, DNA, RNA, protein and lipid content during the postnatal growth of the mouse brain. The three questions below are to be answered by reference to these data.



(1) When does the phase of rapid cell multiplication cease?

1. At birth
2. 7 days
3. 14 days
4. 21 days
5. 28 days

(2) When does the process of myelinization begin?

1. At birth
2. 7 days
3. 14 days
4. 28 days
5. 35 days

(3) Which of the following ratios best expresses the growth due to the average increase of mass per cell?

1. Wet weight/RNA
2. Wet weight/protein
3. Protein/wet weight
4. Protein/lipid
5. Protein/DNA

Graphs, diagrams, and charts of various types can be economically reproduced in large or small quantities to serve as the basis for objective examination questions that compel the student not only to demonstrate that he has acquired the requisite basic information, but also to prove that he can use it to interpret various types of data presented in realistic form and can apply it to the solution of *relevant* problems.

(*b*) *Simulation technique: a new type of test*[1]

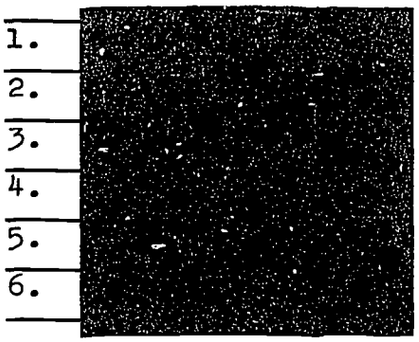
Concern about assessing the student's judgement in solving realistic problems has led to the development of a new type of examination requiring decision making in the solution of laboratory and clinical problems. This new type of test, based on the principles of sequential analysis, utilizes a simulation technique analogous to that employed in business management games and military exercises.

In this type of test, developed jointly by the Center for the Study of Medical Education and the Committee on Student Appraisal of the University of Illinois College of Medicine, each simulated problem in patient management is initiated by a brief verbal description of the patient's chief complaint or by a short colour film in which the patient describes his illness (see Step I of Example 6). The examinee must then decide how he will first approach this patient, i.e., what, if any, action seems indicated at this point. He records this decision by erasing the opaque overlay on a specially constructed answer sheet and finds an instruction directing him to the section designated by his choice (see Step II, Example 6). Here he is confronted with a long list of possible courses of action that will yield further information about the patient (see Step III Example 6). He may select as many or as few

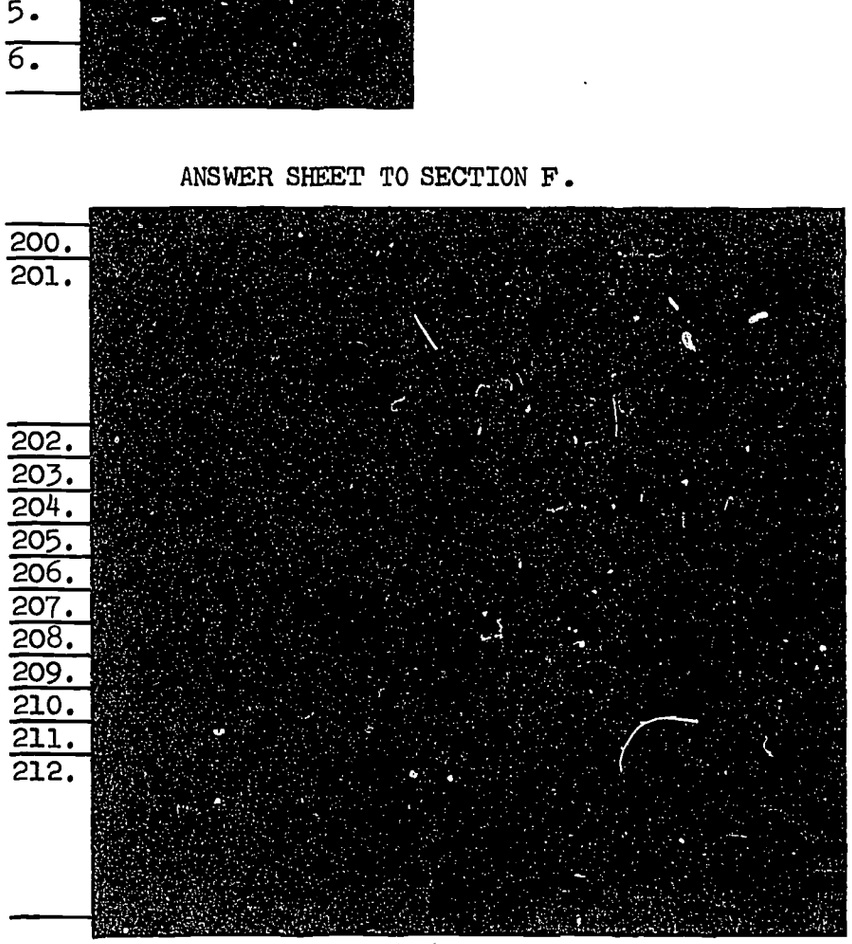
[1] For a fuller explanation of the simulation technique, see Annex 3, p. 58.

INSTRUCTIONS: For each answer, erase the full block, the number of which must correspond to the answer of your choice.

ANSWER SHEET TO PROBLEM



ANSWER SHEET TO SECTION F.



EXAMPLE 6

(Excerpt from a simulated problem in patient management)

Test Booklet	Answer Booklet [1]
Step I	INSTRUCTIONS: For each answer erase the full block, the number of which must correspond to the answer of your choice.
The problem Thirty minutes after a light luncheon a 50-year-old woman executive develops severe abdominal pain during a board of directors meeting. The chairman of the board calls you and asks that you see her as soon as possible. At your request he agrees to arrange for her immediate transfer to a nearby hospital. When you arrive there thirty minutes later you find the patient lying on a trolley in the Emergency Room. She appears to be in severe pain and begs you for relief. Under these circumstances you would FIRST (choose ONLY ONE):	
Step II	
1. Obtain further history	1.
2. Perform a physical examination	2.
3. Initiate laboratory evaluation	3. Turn to Section F
4. Arrange for immediate surgery	4.
5. Arrange for urgent surgery after pre-operative preparation	5.
6. Initiate conservative management without further evaluation	6.
Step III	
Section F	
In light of the available information you would NOW order (select AS MANY AS you consider indicated):	
200. Electrocardiogram	200. See tracing No. 102
201. Complete blood count	201.
202. Blood urea nitrogen determination	202. 12 mg/100 ml
203. CO_2 combining power of blood	203. 25 mEq/l
204. Stool guaiac determination	204.
205. Blood smear	205. See colour plate No. 47
206. Sedimentation rate	206.
207. Haematocrit reading	207.
208. Haemoglobin determination	208. 13.5 g%
209. Barium meal	209.
210. Chest X-ray	210. See X-ray plate No. 72
211. Barium enema	211.
212. Microscopic analysis of urine etc.	212. Bacteria—many Crystals—none Epithelial cells—few Leukocytes—8-10 Red blood cells—1-2 etc. (Magnification per field 10 × 40)

[1] With instructions exposed as though overlay has been erased (see insert).

procedures as seem necessary at that stage, and then, by erasing the appropriate overlays, he finds the "results" of the procedure(s) he has chosen. On the basis of these new data he must decide upon the next step he wishes to take.

Each problem contains many such sections, some of which are not necessarily relevant to the optimal management of the patient. All sections are arranged in scrambled order, and they may be sealed to minimize the possibility of using the options offered in them as clues to the expected choice. In each new section, the examinee must indicate his decisions about a series of specific actions, and at each stage he must make a strategic decision about the overall management of the patient; this decision determines the section to which he is directed next. In this fashion, a problem may be carried through many stages, at each of which the examinee must make further decisions based on the *specific* reactions of the patient evoked by *his own* earlier decisions.

The stages in the management and the responses to the specific procedures the examinee may select are carefully designed to simulate an actual clinical situation. Results of diagnostic and therapeutic procedures are reported in a form resembling the one that the physician customarily encounters. In response to an order for a specific test, a laboratory report is revealed; in response to an order for an X-ray plate, electroencephalogram, electrocardiogram, etc., the examinee is referred to a high-quality photographic reproduction of the X-ray plate or tracing. If he orders a blood smear, he is referred to a colour plate of the smear. If he wishes to obtain auscultatory data, he can be referred to a high-fidelity tape recording. If he orders medication, the patient's response is reported. No interpretation of these data is offered and none is explicitly demanded of the student; he is merely given the data he requests and is required to act on them as does the physician in the conventional clinical setting. However, he may, by making the appropriate erasure, request a consultation for assistance in interpreting the results of any specialized laboratory procedure.

The complications that must be managed differ from student to student, depending (as they do in medicine) on the unique combination of specific procedures each has selected at earlier stages. For some, the erasures will reveal an instruction to by-pass entirely one or more sections of a problem because the approach chosen is effective in avoiding potential complications with which other students would be faced. If, however, at any stage the examinee orders something harmful or fails to take measures essential to the recovery of the patient, he uncovers a description of the clinical features of the complication that has developed. He is then directed to a special section where he has the opportunity to take measures to rectify his previous errors. If these remedial measures

are inadequate, he may be informed that the problem is terminated because the patient has suffered a relapse and has been sent to another hospital, or has been referred to a consultant, or has died.

In general, the problems most amenable to this technique are those involving several stages of data collection and the need for some action after each stage, the propriety of the action depending on data obtained earlier. In addition to the simulated clinical problems in patient management described here, shorter simulation exercises dealing only with one aspect of the diagnostic work-up or with the therapeutic management of a specific patient can be constructed. It should be noted that analogous exercises utilizing simulation techniques have now been developed for laboratory and research problems in the basic sciences.

(c) *Essay and oral examinations: additional designs of new types of tests*

Although the foregoing discussion of new techniques in medical examinations has been devoted primarily to a description of various developments in the use of multiple-choice and other objective examination patterns, it should be noted that these principles apply equally to the use of essay and oral examinations. For example, although judgement and decision-making may be validly measured by objective tests of the new type, such tests are of little value for measuring ability in communicating with one's colleagues or with a patient. Yet, according to the principles of simulation technique the essay or oral examination designed to assess such skills should set a *realistic* task. Thus, a hospital chart for a specific patient could be reproduced and each student instructed to write a discharge letter to the patient's family physician; alternatively, he might be asked to write a referral letter on a patient. Similarly, the oral examination might be used to assess a candidate's ability in taking the patient's history, his skill in examining the patient, or his judgement in determining and defending a plan of management for a specific patient.

Once a decision has been made about the areas in which competence is to be assessed by the oral examination, it is necessary to design *standardized problems* or *situations* in which the candidate will be obliged to demonstrate the level of relevant competence he has achieved. To pursue the previous examples, if the examination is to be used to assess ability in history-taking, then it is possible to standardize such an oral examination by designing (well in advance of the examination) a series of descriptions of the age, educational level, and presenting complaints of a series of patients. At the time of the examination, one or more of these brief descriptions would be given to the candidate and it would be

his task to *interview the examiner* or another person, who, *playing the role* of the patient, would be thoroughly familiar with the "patient's" history. The examiner would then grade the behaviour of the student by observing him.

Similarly, if the oral examination is to be used to test the candidate's skill and knowledge in arriving at a diagnosis and in determining and defending a plan of management, standardized case material may be prepared (again well in advance of the examination).[1] This material would be presented to the candidate at the time of the examination. It would be his task to discuss with the examiner his diagnostic impressions, his reasons for them, and the next steps he would recommend for the care of a specific patient, and to defend his therapeutic decisions. Alternatively, the student could be given an X-ray plate and asked to reconstruct the history that led to the findings and to outline and justify his recommendations for management. Thus, his skill in reading radiographs, his ability to observe and to make a synthesis of available data, as well as his therapeutic judgement, could be assessed simultaneously.

Once a decision has been made on what the essay or oral examination is to measure and standardized test situations have been developed to accomplish this purpose, it is imperative to decide what criteria and standards shall be used in assessing the candidate's performance and to make certain that these are uniformly applied. To this end, it would be desirable to isolate and specify in concrete behavioural terms the various factors that distinguish competence from incompetence.

Finally, these illustrations suggest that appropriate use of the oral examination requires a substantial revision of the examiner's role and responsibilities. Instead of merely acting as a "quiz-master", he must assume responsibility at all stages of the examining process. In advance of the examination he must make certain policy decisions regarding what is to be assessed and he must design examination problems that are compatible with those policy decisions. At the time of the examination he may be required to play various roles, from simulating a patient to merely observing a candidate. However, it must be mentioned that the examination technique in which the examiner or another person plays the role of the patient is subject to some criticism. An examiner who plays the role of the patient may well represent almost as much a variable as an actual patient, and the design of such techniques may not be significantly less difficult than finding representative patients. It is therefore proposed that such techniques should be used mainly as a

[1] This, however, describes an ideal situation and the average teaching hospital often cannot keep patients available over a sufficiently long period. Even to secure out-patients for such examinations may at times be a difficult and unreliable procedure.

preliminary step to be followed by the examination of the patient himself (see below) and that the student's competence be assessed accordingly. It is hoped that well designed simulation techniques in preclinical subjects will become one of the universally applied methods of examination, although the present high costs of preparing the relevant audiovisual material will, for many a school, limit the use of these examinations in the near future. Nevertheless, whatever method of grading is used, some examiners will need to learn how to define and apply uniformly a set of predetermined standards. This shift in the role of the examiner may require his further training in order to assist him in learning how to discharge his new functions.[1]

(d) The practical examination

It is often argued that exercises in the interpretation of data, and even the written and oral simulation tests described above, are not realistic because they only *simulate* but do not *duplicate* the "real-life" situation. In an effort to make the assessment of professional competence more relevant and more valid, a number of specialty boards and some medical schools employ a "practical" examination in which, for example, the candidate is assigned a patient and is *observed* by the examiner while taking the history and performing the physical examination. During this observation the examiner rates the candidate on his skill in eliciting clinical information and on the accuracy of his findings. The examiner then discusses with the candidate the latter's recommendations for the next steps in the management of this specific patient and makes an additional rating of the quality of clinical judgement revealed in this discussion. This type of examination certainly appears to meet satisfactory standards of relevance and validity since it yields an actual sample, in a realistic context, of the types of skill the physician must display daily.

Considering the foregoing discussion, however, the practical as well as the more common type of oral examination has serious shortcomings with respect to the reliability of the information it yields. Systematic analyses generally reveal a far greater range of "examiner variation" with regard to standards and mode of procedure than is commonly recognized. This examiner variation can be reduced if a committee decides on how to apply a given examination routine.

Further, and even under the best of circumstances, it is obvious that both oral and practical examinations are very time-consuming, and therefore the sample of candidate behaviour that can be obtained within

[1] See " Newer developments in the training of examiners ", page 48.

a given period is necessarily smaller than in a carefully constructed written objective examination. This restriction of sample size in itself limits the reliability of such examinations. Their reliability is further reduced by the fact that the examination situation is usually not standardized; it often varies significantly from candidate to candidate, depending not only on the factors listed in the preceding paragraph but also on the patient assigned to the candidate and on the candidate's familiarity with the subject first discussed, his knowledge of the examiner's special interests, or his skill in "leading" the examiner. Candidates are well aware of these variations and, when they know the identity of their examiners in advance, it is common practice for them to "study the examiner" rather than the subject. Clearly, these factors reduce the reliability and thus the validity the practical examination would otherwise possess.

For these reasons a "simulation technique" employing *standardized* cases, *standardized* situations and *standardized* criteria for judging candidate competence is now being applied whenever possible in place of the older style of oral and practical bedside or laboratory examinations. This new approach should not exclude the careful selection of patients provided the standardization of assessment of student performance is agreed upon.

Therefore, schools that have adopted the newer approach to testing use the practical examination only if it clearly constitutes the most feasible and valid measure of a relevant component of the student's performance. Schools that experience over the years a great variation in the pass-fail ratio[1] should make every effort to increase reliability of grading. This can be achieved by (*a*) selecting a limited aspect of what the student is expected to know, (*b*) standardizing the test situation used to assess it, and (*c*) developing an objective checklist for the examiner to use in grading. To standardize the test situation, one might, for example, design a practical examination in microbiology in which each student is observed as he takes a throat culture and prepares the smear. In a clinical discipline a practical examination might consist of observing each student as he performs a neurologic examination on an individual who has been carefully trained to simulate a particular neurologic disease. In the latter case, a checklist such as that shown in Example 7 would greatly improve the reliability of the observations made by the examiner. Employing such a check-list has three distinct advantages: (1) it greatly increases agreement between examiners as to the quality of the student's performance; (2) it provides a better feedback of information to the

[1] In some schools in the USA the pass-fail ratio may be as high as 30: 1, whereas in some schools in Europe it may be only 3: 1, but in any given school it should remain fairly steady.

EXAMPLE 7

Student's name: ..... Type of patient: .....

Instructions to examiner:

For each item listed below, check the appropriate column.

	Yes	No	Not relevant
1. Did the student examine the optic fundi?	.....	.....	.....
2. Did the student examine the reflexes?	.....	.....	.....
3. Did the student examine the chest? etc.	.....	.....	.....
10. Did the student adequately expose the area to be examined?	.....	.....	.....
11. Did he perform the examination with minimal discomfort to the patient?	.....	.....	.....
12. Did he show concern for the patient's physical condition and sensibilities? etc.	.....	.....	.....

	Outstanding	Satisfactory	Not satisfactory
20. What is your *overall* evaluation of the candidate's performance:	.....	.....	.....

Comments: .....

.....

.....

.....

Examiner: ..... Date: .....

student, who can then identify and correct his deficiencies; and (3) it enables the teaching faculty to identify specific strengths and weaknesses that are relatively common among the students they have taught and thus provides a better indication of desirable revisions in the teaching programme.

3. *Development of appropriate tests in the affective domain* (*habits, attitudes, values*)

It has been found that the written, oral and practical simulation techniques described above yield some data about a student's attitudes and habits. For example, in the written examinations some students

repeatedly initiate laboratory investigation or therapy without adequate inquiry into the diagnosis, while others repeatedly fail to take decisive action even when the need for it has been clearly established. Similarly, gross differences in the consideration students show for a patient's comfort and welfare may become obvious in the practical examinations. However, one cannot be sure that the attitudes shown by a student in an examination are typical of his habitual performance—the sample of behaviour is too small and bias may be introduced by the conditions of the examination. For this reason, the newer approach to the assessment of professional habits, attitudes and values places heavy emphasis on obtaining *descriptive* reports from many instructors who know the student well because they have had the opportunity to observe him in many types of setting over a long period of time. With a view to enhancing the reliability and objectivity of these assessments of attitude, two methods are now being introduced. One calls for objective, anecdotal statements by the observer (i.e., instructor), briefly describing the setting and the specific action of a student which he regards as "outstanding" (i.e., as evidence of either superior or unsatisfactory professional behaviour). The accumulation of such statements during the student's medical school career serves to identify the major qualities characteristic of the student. The second approach to the assessment of habits and attitudes is to identify critical variables in professional behaviour, to provide descriptive statements of different types of behaviour in regard to each, and to ask the student's instructors to check the one that best characterizes the student's usual behaviour with respect to each variable (see Example 8).

EXAMPLE 8

Variable: Response to criticism (check one):

1.	2.	3.	4.	5.
Accepts criticism easily; makes you feel he appreciates your interest in his shortcomings.	Accepts criticism and asks pertinent questions about the matter under discussion.	Accepts criticism stoically.	Does *not* accept criticism well; presents various excuses to explain his shortcomings.	Becomes silent, resentful, or overtly hostile when criticized.

It is argued that approaches of this kind have three major advantages: (1) they yield more objective and reliable information; (2) they enable the student to identify more precisely weaknesses in his medical know-

ledge and help the school to pinpoint areas in which it may have been negligent in emphasizing appropriate models of professional behaviour; and (3) they help to avoid the confusion produced by a single overall assessment in which, for example, students who demonstrate superior intellectual abilities but unsatisfactory professional attitudes are given the same "grade" as those whose intellectual performance is weak but who attempt to compensate by demonstrating superior professional attitudes.

D. NEW APPROACHES TO THE REPORTING AND ANALYSIS OF EXAMINATION DATA

The advances made during the last 20 years in educational psychology exert increasing influence on the thinking of the faculty staff in medical schools. A few medical schools now have their own departments of medical education, but in others there is no formally constituted department for organizing clinical studies. However, at some universities there are educational departments or teacher training centres on the same campus as medical schools, and this tends to encourage joint action on problems of medical education (see page 49). Some of the ways in which examination results may be used to help both the student and the faculty are discussed below.

1. *Reports to the student.* Reports of examination results nowadays more and more frequently provide the student with a "profile" describing, in both absolute terms and in relative ones (i.e., by comparison with others in his group), his strengths and weaknesses with respect to subject matter as well as to intellectual (or other) attributes. When this policy is followed, the student receives, in addition to an overall grade, a detailed analysis of his performance which (1) indicates his scores on questions classified either according to subject areas (infectious diseases, cardiac problems, body fluid metabolism, etc.), following an organ-system teaching approach, or by faculty departments (microbiology, internal medicine, paediatrics, etc.), following a non-integrated teaching approach; and (2) reports his scores on the same questions classified according to the requisite performance such as recall, problem-solving, clinical judgement, and skill in performing physical examinations.

2. *Reports to the faculty.* It is becoming increasingly common not merely to report the number of failures but to provide individual departments and the faculty as a whole with a "group profile" of their students, which includes a detailed analysis of the number who performed satisfactorily with respect to each subject group and with respect to each type of intellectual and other skill measured by the test (see above).

With increasing frequency, the relevant department is also provided with a detailed report on the number of students who gave *each* answer to *each* question, so that the teaching staff can determine which facts, concepts and approaches have been found to be well understood and what misconceptions are still prevalent.

Armed with these two types of report, a departmental staff can then make rational changes in its curricular programme and instructional techniques.

3. *Reports to the examiners.* Finally, the practice of reporting to examiners on the quality of the examination itself is becoming increasingly widespread. Such reports normally give information about the objectivity, reliability and validity of an examination and its relation to other measures of student competence, and they furnish a detailed analysis of the level of difficulty and discrimination of each question, together with an indication of questions that are technically deficient and suggestions for improving them.

4. *Use of examination data.* It is apparent that this newer system of reporting examination is of considerable importance in that (1) it can help to raise the quality of both instruction and evaluation; (2) it can encourage the individual student (and his instructors) to individualize learning and so make it more effective; and (3) it enables promotion committees and other certifying bodies to make more rational decisions about the careers of individual students.

E. NEW APPROACHES TO THE PROBLEM OF SETTING STANDARDS OF COMPETENCE

Once a test has been developed to assess the various abilities that go to make up clinical competence there remains the difficult problem of determining what degree of competence is enough to warrant certification or licensure. Despite long-standing practices throughout the educational system, this simply cannot be decided by the use of some arbitrary numerical symbol such as 75%, B+, or 12/20. It must be evident that "75%" on one test could represent outstanding performance, while the same grade on another test covering the same content would be clearly unsatisfactory, *depending on the difficulty of the questions asked.*

Basically, there are only two methods of evaluating an individual's performance in any achievement test, including examinations for licensure:

(1) by judging his performance *in relation to that of his group;* or

(2) by judging his performance *in terms of an absolute standard.*

First method: It is clear that no matter how "the group" may be defined or what particular statistical manipulations may be employed (for example, inspecting the distribution for "breaks" or converting raw scores to standard scores or scaled scores), this is, in essence, "grading on the curve". This system is widely used in the USA for the standardized tests given in elementary and even secondary schools, where the reference group is composed of a large, representative, nation-wide sample collected over a period of several years. Even under optimal circumstances this method has the drawback that standards are set in terms of what "is" rather than in terms of what "ought to be". It also means that an individual may "look good" either because he achieves a great deal or because he simply has the "good fortune" to be a member of a group most of whom achieve very little. (When relative standards are applied in examinations for licensure, some physicians who are certified as satisfactory in one year would not necessarily have been so certified had they taken the examination with a different group a year earlier or a year later.) This first method is particularly deficient if the reference group comes from a highly selected population (such as physicians) and is composed of those few who happen to take the examination at the same time and place. A third serious shortcoming is that the group itself becomes the arbiter of the standards by which it is to be judged.

Second method: Judging the individual's performance in terms of a criterion of adequacy that is independent of the performance of the particular group of which he is a member has also certain obvious drawbacks: (1) people do not usually agree about standards; (2) even if agreement on general standards can be reached, there remains the difficulty of obtaining specific agreement about the level of performance required in a particular test to meet the generally accepted standard; (3) expectations may need to be revised in the light of experience with a particular type of test; (4) finally, compromises in standards may be required in order to bring them into accord with the reality of a particular learning situation (for example, it may be desirable but "unrealistic" to require that all candidates for licensure should be familiar with clinical research techniques, if that goal is assigned a lower priority than the attainment of skill in the examination, diagnosis and therapy of the patient and if the time available is inadequate).

This second method does, however, have the effect of giving examiners the power as well as the responsibility of setting standards of competence. Despite its inherent difficulties, it is uniquely suited to the evaluation of individuals who come from a highly selected population and who belong to a single community of scholars having a common concept, however vaguely defined, of what constitutes acceptable

technical and behavioural standards of the profession. As applied to an examination for licensure, the method requires that the standard be set in terms of a level of competence, not in terms of relative position in a group. Furthermore, the determination of this level of competence should reflect a detailed judgement about the specific character of individual questions and problems, not merely a general estimate of what a candidate "ought to do on tests like this".

Finally, if a candidate is to be tested for professional competence, then he should be judged on whether he possesses the required degree of "competence", whatever it may be, and not on the means by which he acquired it or the time it took him to achieve it. In practical terms, this means that the completion of specific courses, or the rigid requirement of a fixed number of months or years in a particular kind of educational programme, is not a justifiable criterion for judging professional behaviour.

This approach to the setting of standards has three distinct advantages: (1) it avoids fluctuations in standards which would harm the community either by permitting inadequate students to be certified or by depriving the community of the services of satisfactory ones; (2) it enables the examiners to identify changes in the performance of the students by significant changes in the failure rate; and (3) without any lowering of standards, it tends to avoid unwarranted rigidity in medical education and improves the prospect of more efficient and effective manpower training and utilization—a matter of the highest priority in the health professions.

F. NEWER DEVELOPMENTS IN THE TRAINING OF EXAMINERS

As noted in the section "Essay and oral examinations" above, the introduction of these newer techniques of analysing, constructing and evaluating examinations and their results shifts the responsibilities of the examiner and requires that he develop new knowledge and new skills. This urgent necessity was recognized by the WHO Expert Committee on Professional and Technical Education of Medical and Auxiliary Personnel, who expressed it as follows:[1]

> The teachers need help in acquiring the skill that will allow them to design new test procedures or vary old ones and the understanding that will allow them to score accurately, interpret perceptively, report meaningfully and use wisely the information derived from the measurement methods they use.

[1] *Wld Hlth Org. techn. Rep. Ser.*, 1966, No. 337, p. 10.

Opportunities to obtain this training are becoming increasingly available. In Europe and the USA there exist departments of education attached to many universities or established as independent institutions. Medical schools often do not know or make use of these departments although their experience in educational psychology could be shared with the staff of any medical school. A department of education could, in co-operation with a medical school, start on research projects in medical education. Such joint action would eventually have an impact on faculty staff attitudes towards teaching, and the following activities could be a start in this direction: (1) short (1-2-day) meetings of faculty or interfaculty committees who, with the assistance of expert consultants, learn how to develop and to apply technical criteria to the analysis of existing examination procedures and to the specification of abilities to be assessed; (2) short meetings of faculty or interfaculty task forces who, with the assistance of expert consultants, learn how to construct new types of examinations and begin to develop a pool of more reliable and valid test materials; (3) practice sessions designed to assist a cadre of special examiners who have the responsibility of administering a specific oral or practical examination of the new type; (4) short training courses for faculty or interfaculty task forces charged with the review of examination results relevant to curricular and other policy decisions, the purpose being to assure appropriate exploitation and interpretation of these results; (5) more extended (1-6-week) formal training programmes in educational research design and evaluation for interested individuals; and (6) formal fellowships in research in medical education, lasting up to one or two years, and sometimes leading to a postgraduate degree.

However, especially with regard to the last two points, when and where such programmes are set up is often left to chance, information about them is limited, and their number is as yet far from adequate. The success of these programmes often depends on the extent to which the department of education is aware of the problems and setting of medical education. For this reason, it is desirable to establish as a regional centre a separate division for training and research in medical education within a medical school. The influence of such a centre on other medical faculties in the region will eventually become apparent. The academic board, or governing body, should take an interest in research in medical education so that changes proposed on the basis of systematic evidence can be decided upon at the highest level. WHO could render an extremely valuable service if it not only acted as a clearing house for information about such programmes, but also gave direct support and encouragement to their expansion.

Annex 1

A POSSIBLE INTERNATIONAL QUALIFYING EXAMINATION FOR MEDICAL STUDENTS

The report of a study group on internationally acceptable minimum standards of medical education[1] suggested two possible approaches to achieving equivalence at the international level in the appraisal of medical training: the accreditation of schools, or the introduction of some sort of international qualifying examination. The latter should be designed for use towards the end of the regular university programme in medicine, just before the students begin formal apprenticeship as interns, house officers, junior physicians, etc. in a hospital. We believe that a pilot study should be carried out prior to any decisions about an international qualifying examination.

The examination used in such a pilot study should test relevant aspects of physiology, pathology and biochemistry, as well as a fairly narrow range of clinical subjects possibly limited to internal medicine, paediatrics or gynaecology and obstetrics; it should also include a sample of the major medical or surgical emergencies that a physician at this level of training could be expected to meet. We recognize that emphasis on different areas of medicine (e.g., infectious diseases, geriatric medicine, tropical medicine, chronic diseases) varies throughout the world. However, these variations could be taken into account by careful design and appropriate sampling techniques in the examination. The examination should be designed to assess various levels of understanding: the ability to recall pertinent information, the ability to observe and interpret data, and the ability to work through problems in patient management.

Finally, if this pilot study were conducted in a number of countries with sufficiently large samples of students at comparable levels of instruction, it would furnish enough information to point out some major differences between countries and thus indicate those where further investigation would be required before a comprehensive international examination could be designed.

---

[1] *Wld Hlth Org. techn. Rep. Ser.*, 1962, No. 239.

Annex 2

AN ILLUSTRATIVE LIST OF CRITICAL PERFORMANCE REQUIREMENTS FOR PHYSICIANS

A. *Cognitive domain*

1. *Knowledge* of fundamental technical vocabulary, facts, concepts, principles, laws, methods, and procedures as demonstrated by:

(*a*) accurate recall;
(*b*) accurate recognition.

2. *Understanding* of these facts, concepts, etc., as demonstrated by: the ability to:

(*a*) explain them;
(*b*) recognize their implications;
(*c*) use them for the explanation of phenomena.

3. *Ability to analyse and interpret data* of various types as demonstrated by:

(*a*) accurate translation from one form to another;
(*b*) formulation of plausible hypotheses to explain data;
(*c*) formulation of plausible predictions;
(*d*) recognition of limitations of data.

4. *Ability to solve relevant problems*, as demonstrated by:

(*a*) recognition of the data required to solve the problem;
(*b*) utilization of appropriate sources to obtain the required data (e.g., selecting or ordering appropriate X-ray photographs or laboratory tests);
(*c*) formulation of a tentative hypothesis (or diagnosis);

(*d*) recognition of appropriate methods for checking the hypothesis (or diagnosis);
(*e*) formulation of a plausible scheme of therapy.

5. *Ability to take a history*, as demonstrated by:

(*a*) eliciting the chief complaint;
(*b*) obtaining a clear description of the present illness;
(*c*) following up positive leads in the history;
(*d*) obtaining adequate information about past illnesses and family history;
(*e*) obtaining adequate information about each system;
(*f*) using vocabulary and form of inquiry appropriate to the patient's comprehension and co-operation;

6. *Ability to retrieve information and to keep records.*

7. *Ability to utilize community resources.*

8. *Judgement in evaluating a complex situation*, such as research, laboratory, clinical or community problems, when for example:

(*a*) dealing with complicated clinical cases by:

(i) recognition of the urgency or seriousness of the situation;
(ii) adjustment of the nature of the history-taking and physical examination to the requirements of the specific situation;
(iii) recognition of the need for special additional diagnostic methods, such as repeated X-ray examinations or laboratory determinations, and interpretation of these findings;

(*b*) establishing a correct diagnosis in complicated cases by:

(i) double-checking of unexpected findings;
(ii) persisting till a definitive diagnosis has been established;
(iii) recognition of the primary disorder;
(iv) recognition of underlying or associated problems;
(v) adequate care to rule out other disorders, etc.;

(*c*) making the right decisions for ordering appropriate management in complicated cases by:

(i) determination of kind, extent and immediacy of needs;
(ii) planning the patient management for a given situation;

(iii) adaptation of treatment plans to the individual patient with due consideration for patient's age, general health, special needs, or a specific condition that may require special attention to therapeutic contraindications;

(iv) checking the effectiveness of therapy by monitoring the patient's progress;

(v) reassessment and modification of treatment plans in response to changes in patient's condition;

(vi) arrangements for follow-up and long-term care, including appropriate use of referral services for physical, social and economic rehabilitation;

B. *Psychomotor domain*

1. *Skill in performing physical examinations*, as demonstrated by:

(*a*) performance of a thorough general examination;
(*b*) accurate detection of all significant physical signs by inspection, percussion, palpation and auscultation;
(*c*) performance of examination without causing the patient undue pain or embarrassment.

2. *Skill in using various laboratory and clinical instruments*, e.g., the microscope or ophthalmoscope.

3. *Skill in performing technical procedures*, such as venepunctures, lumbar puncture, catheterization, intubation, preparing a specimen, or handling delicate biological materials.

C. *Affective domain*

1. *Concern for patient and patient's family*, as demonstrated by:

(*a*) a personal interest in, and acceptance of responsibility for, the patient's welfare;
(*b*) a discreet and tactful manner when dealing with the patient and his family;
(*c*) awareness of the patient's anxiety, which should be allayed by reassurance and support;
(*d*) frank discussion with the patient and family to explain his condition, treatment, prognosis or potential complications.

2. *Awareness of his own professional capabilities and limitations*, in particular :

(*a*) acts only within his own area of competence, unless forced by an emergency to help in another specialty;
(*b*) admits areas of ignorance and error;
(*c*) seeks help, advice or consultation.

3. *Willingness to establish effective relationship with colleagues and other members of the health team*, and to:

(*a*) accept suggestions and criticism;
(*b*) handle differences of opinion discreetly and tactfully;
(*c*) give support and direction to less experienced personnel;
(*d*) take responsibility for his own decisions.

4. *Willingness to develop and to apply an inquiring mind* in order to:

(*a*) reconsider cherished convictions;
(*b*) actively seek new knowledge.

5. *Organization and utilization of own specialized knowledge and skills* to contribute to community as well as to individual patient welfare.

The above items A. and C. of critical performance requirements for physicians are based on *Taxonomy of Educational Objectives*, Handbooks I and II.[1]

---

[1] Bloom, B. S. & Krathwohl, D. (1956) *Taxonomy of educational objectives. Handbook I: Cognitive domain*, New York, McKay; Krathwohl, D., Bloom, B. S. & Masia, B. (1964) *Taxonomy of educational objectives, Handbook II: Affective domain*, New York, McKay.

Annex 3

EXAMINATION METHODS

IMPROVING PRESENT EXAMINATIONS

A review of the examinations at present being used in Europe and North America reveals the following principal shortcomings: (*a*) triviality of the questions asked; (*b*) unintended cues to the correct answer in the formulation of questions; (*c*) outright errors in phrasing of questions (or, in the case of multiple-choice examinations, in phrasing of the correct response); (*d*) ambiguity or complexity of language; (*e*) forcing the student to answer in terms of the provincialism or personal views of the examiner. It follows that the safeguards to be introduced for the improvement of examinations can be summarized under three main headings: (1) adequacy and accuracy of the questions from the point of view of subject matter; (2) adequacy of the questions from a technical point of view; and (3) adequacy of the questions from the point of view of the type of competence the test purports to assess.

Subject matter considerations

Analysis of both oral and written tests indicates that some and often many of the questions utilized are ambiguous, unclear, controversial, esoteric or trivial. The author of both essay and objective questions should always submit them to the critical review of his colleagues in order to assure (*a*) that the content being sampled is of general importance and not merely a matter of special interest to the author; (*b*) that the content is relevant to general practice or to specialties other than the author's; and (*c*) that the questions (and the answers in the case of multiple-choice examinations) are so formulated that agreement can be reached on what constitutes an appropriate answer. Finally, it is clear that such a review would help to avoid the oversimplification characteristic of many tests, which so often leads to the conclusion that "the more you know about the subject the lower your score will be".

Technical considerations

The technically adequate question has certain essential characteristics.

First, the instructions and the question that is being asked must be clear and unequivocal. While this requirement seems obvious, questions frequently fail to meet it. This failure is common to both objective and essay questions. In the case of the former, directions are frequently obscure, inadequate, confusing, or unnecessarily complex, and the essay question is often so vaguely formulated as to permit students (consciously or not) to evade the real issue. Frequently, both essay and objective questions seem to be saying: "Guess what I (the author) am thinking about". The probability of an appropriate response should *not* depend on knowing who wrote the question (or even which department submitted it) so that "giving the answer that he wants" becomes easier.

Second, and this is particularly applicable to the objective type of test, questions must be worded so as to avoid giving clues that assist the uninformed student in answering correctly without really understanding the subject. For example, if the students know that the correct answer is likely to be longer than the others, formulated in very technical language, and hedged round with qualifications, they will choose it without necessarily understanding the content. In general, it is also wise to avoid extreme statements as students reject these as a matter of principle. In short, the correct and incorrect answers in a multiple-choice question should be of similar length, type, and technical intricacy.

Third, it is especially important to ensure that the correct and incorrect answers in multiple-choice tests have the same grammatical form, for the student may otherwise reject the incorrect answers simply because they do not agree grammatically with the question asked.

Fourth, it is usually preferable that both correct and incorrect answers to multiple-choice questions refer to the same general subject. Questions in which one answer deals with the etiology, another with the symptoms, and another with the therapy of a particular syndrome, for example, had best be avoided. Such questions would ordinarily be improved if divided into three separate multiple-choice questions, the first containing a correct statement and four or five incorrect statements about etiology, the second referring to symptom complexes, and the third to therapy.

Fifth, in constructing the "wrong answers" for a multiple-choice question, it is important to utilize common, plausible misconceptions instead of relying on "trick" formulations, silly and outrageous statements, or contrived misinterpretations. The incompetent student is most likely to demonstrate his ignorance when the wrong answers to

a question represent intelligent formulations of the confusions and misconceptions to which he is actually subject.

Sixth, in general, objective questions that instruct the student to select the *best one* of several responses are superior to questions that instruct him to check *all* the correct responses or to judge the truth or falsity of isolated statements. It is difficult (perhaps impossible) to write discriminating questions based on the assumption that experts will agree that any single, simple statement is clearly correct or wholly true.

Seventh, while extraneous material may sometimes be interesting, it is rarely helpful and may actually be distracting in a question. For example, the simple question "Antinuclear antibody is found in which of the following?" is far more effective than "Antinuclear antibody actually includes a family of antibodies against various nuclear constituents. It is found in virtually all cases of systemic lupus erythematosus. It is also found in which of the following?"

Eighth, the group of questions comprising a test must constitute an adequate, representative, and appropriate sample of the course content and of the intellectual behaviour it is designed to promote in students. Clearly, if the questions are too few in number or are unbalanced in emphasis, the test scores will be unreliable and misleading. Similarly, if the difficulty or complexity of the questions is inappropriate, the test will not be effective in discriminating among students. If they are too easy, it will be impossible to separate the mediocre from the excellent students; if too difficult, it will be impossible to distinguish the incompetent from the barely adequate and the mediocre students.

Types of competence a test purports to measure

No one kind of test (objective, essay or oral) is superior to all others for the measurement of the higher and more complex intellectual processes. Studies of various types of test support the view that the essay and the oral examination, as commonly employed in medical schools, test predominantly simple recall and, like the objective tests in current use, rarely require the student to engage in reasoning and problem-solving. In short, the *form* of a question does *not* determine the nature of the intellectual process required to answer it.

Second, there is often a tendency to confuse the difficulty of a question with the complexity of the intellectual process measured by it. However, a question requiring simple recall may be very "difficult" because of the esoteric nature of the information demanded; alternatively, a question requiring interpretation of data or application of principles may be quite "easy" because the principles of interpretation are so familiar and the

data to be analysed are so simple. In short, the difficulty of a question and the complexity of instructions are not necessarily related to the nature of the intellectual process being tested.

Third, there is often a strong inclination to assume that any question that includes data about a specific case involves problem-solving. In fact, the "data" are often merely "window dressing" when the question refers to a general condition and can be answered equally well without reference to the data. In other questions, the data furnished about a "specific case" may constitute a "cut-and-dried", classical textbook picture which simply requires the student to recall the usual symptoms associated with a specific diagnosis. Questions of this type can be readily converted into problems that do require interpretation and evaluation of data simply by making the case material conform more closely to that presented by an actual patient than to a textbook description.

In sum, a test that purports to measure the student's clinical judgement and his ability to solve clinical problems must simulate reality as closely as possible by presenting him with constellations of data that are in some respects unique and, in that sense, new to him (see discussion of "simulation techniques" on page 36 and below).

THE DEVELOPMENT OF SIMULATION EXERCISES IN MEDICINE

A clinical problem that purports to simulate the physician-patient encounter must have the following characteristics: first, it must be introduced by information of the type a patient gives a physician, not by a predigested summary of the salient features of the case, and if it is to be realistic it must be described in terms that the patient or a referring physician would use. Second, the exercise must require a series of sequential, interdependent decisions representing the various stages in reaching a diagnosis and in management of the patient. Third, the examinee must be able to obtain in realistic form information about the results of each decision, as a basis for subsequent action, Fourth, once these data have been obtained it must be impossible for the examinee to retract a decision that is revealed to be ineffectual or harmful. Fifth, the problem must be constructed so as to allow both for different medical approaches and for variation in patient responses appropriate to these several approaches. Accordingly, provision must be made for modifications in the problem as the patient responds to the specific courses of action chosen by each examinee.

In the selection of problem for development, it is essential to avoid a uniform, stereotyped pattern that would have the effect of rewarding the same general type of approach throughout. For example, some

problems may deal with emergency situations in which the initial diagnostic enquiries should be kept to a minimum; here the inexperienced medical student may erroneously subject the patient to stressful diagnostic procedures or may hesitate far too long before instituting emergency treatment. In contrast, other problems may involve conditions that require thorough evaluation prior to any decision about therapy; here the inexperienced student may fail to withhold treatment long enough to collect all the data essential to intelligent management. In other problems that are diagnostically simple, the student may be tempted to indulge in over-elaborate investigations, continuing to order tests long after he has obtained adequate confirmation of his diagnosis. Other exercises may deal with the long-term course of a chronic disease. Finally, any diagnostic or therapeutic procedure carries a potential risk for some patients, and thus, in any problem, unique iatrogenic complications may become a major source of difficulty for the student.

In developing each section of a problem, it is necessary, as in other types of test, to avoid providing clues that are artifacts of the test technique. In simulation exercises this means that each section must offer numerous possible interventions, apparently a random sample of the medical arsenal of diagnostic and therapeutic methods. While each section must give the *appearance* of a random listing, it must *in fact* offer a carefully structured group of procedures that not only permit the student to obtain the information he needs for successful handling of the problem, but also provide ample opportunity at every stage to pursue any of the commonly held, plausible but erroneous hypotheses. Finally, it is essential that few if any data be gratuitously provided, so that all decisions, even those that appear most routine, become the responsibility of the examinee. When scoring simulation exercises, a group of experts in the relevant specialty assigns each of the several hundred choices available in a problem to one of five categories, ranging from "clearly contraindicated" to "clearly indicated and important" in the care of *this patient*, at *this time*, under *the conditions specified.* Each choice is then accorded a positive or negative weight of a magnitude that reflects the judgement of the expert group. In this way, it is possible to assign quite objectively to each student a "proficiency score", which represents the degree of agreement between his choices and those of the expert group, and thus to identify numerically the various combinations of choices that constitute skilled management and those that represent merely adequate, or even totally inadequate, care of the patient.

Finally, various patterns of proficiency and error scores (both errors of omission and errors of commission) that reflect different problem-solving styles and approaches can be identified. For example, the decisions of some examinees correspond closely to those of the expert

group (i.e., they select most of the measures that the expert group regards as clearly indicated and avoid most of those classified as contraindicated); these examinees may be described as both thorough and discriminating in their approach. Others, both students and clinicians, make a combination of choices that can only be characterized as a "shotgun" approach to patient care. These examinees and clinicians have moderate to low proficiency scores, usually combined with many errors of commission and few errors of omission. Other examinees make a combination of choices that can best be described as a constricted approach to medical problem. They, too, have moderate to low proficiency scores with few errors of commission and many errors of omission. These data can be used to assist the student in improving his approach to clinical (and laboratory) problems.

Annex 4

EXAMINATION PRACTICES IN SELECTED COUNTRIES

CANADA

Selection of students

In October 1966 an agreement was reached between all Canadian medical schools to apply the Medical College Admissions Test (MCAT) for the selection of students. French-speaking schools are supplied with the appropriate translation of the test.

Assessment of student performance

In general, formal examinations "for the record" are administered only at the end of each course of study and are the responsibility of autonomous departments. To date there have been no interdepartmental examinations except as informally arranged by co-operating departments within a school. However, one school reports the appointment of a new examination committee whose duty it will be to develop, administer and grade all examinations in the clinical disciplines.

Heavy reliance is placed on written examinations of the essay type in assessing Canadian medical students. In a study carried out in 1965, analysis of the examinations of the four western medical schools for the previous 5 years revealed that about 80% of all evaluations (cognitive) were based on essay-type examinations. This was true of all disciplines—basic science as well as clinical. Skills were appraised by laboratory procedures in basic science and by bedside examinations in the clinical years. In some areas these were supplemented by an oral examination. Analysing these traditional medical school examinations in the western provinces, Gilbert[1] found that, as a rough approximation, 95% of the questions involved information recall, 5% generalization, and virtually none were of the problem-solving type.

External examiners play a minor role in Canadian examinations. Their use has been infrequent and irregular, though some departments in some schools practise an informal exchange with their counterparts

[1] Gilbert, J. A. L. (1966). In: *Medical Meetings: The Twenty-fourth Annual Meeting of the Association of Canadian Medical Colleges, Canad. med. Ass. J.*, 95, 983.

in other schools, and some departments make use of specialized National Board examinations (see USA, page 74).

Licensure of physicians

The Medical Council for Canada (MCC) was set up in 1912 as an examining board for certification. Unlike the National Boards in the USA, it deals only with clinical subjects—medicine, surgery, psychiatry, preventive medicine, obstetrics and gynaecology, and paediatrics. The examination (LMCC) consists of a five-question essay-type paper in each of the above subjects, followed by an oral examination in preventive medicine and both a bedside and an oral examination in the other clinical disciplines.

Until recently modified, the written examinations of the Medical Council of Canada tested predominantly (85%) the ability to recall. In the oral examinations, approximately 60% of the questions demand little more than recall of information in the major disciplines; the same is true of nearly 95% of the questions in the minor disciplines. The other elements in the oral examinations are problem-solving (in approximately 18% of the questions) and evaluation of a total situation (in approximately 13%).[1]

The examination is held twice a year for all medical schools in Canada. The purpose of the examination, like that of Part III of the National Board examination (see USA), is to evaluate the candidate's fitness to practise. Though the certificate is not granted until the completion of a satisfactory internship, both the written and the oral examination can be taken at any time following the completion of the medical course. In practice, therefore, many schools use the LMCC written examinations for their final M.D. examinations in medicine, surgery, etc., grade them locally for medical school purposes, and then send them to Ottawa for LMCC to grade for licensure purposes. In this way, the one licensure examination is made to serve two purposes, and thus the number of written examinations is reduced. In addition, the student must take two practical examinations, one for the school and one for the LMCC.

CZECHOSLOVAKIA

During the past twenty years the system of instruction and examination in the medical schools of Czechoslovakia has been reformed several times. All teaching and evaluation activities are under the critical supervision of a "pedagogical committee" convened by the dean's

[1] Gilbert, J. A. L. (see footnote on preceding page).

office. The chairman of the committee is one of the vice-deans and the members are selected from among the experienced teachers, irrespective of their rank. The committee sits at least once a month, and decides directly on minor issues. More important problems are brought before the monthly plenary session of the faculty.

Selection of students

Satisfactory grades in the final examination *(maturitní zkouška)* of the secondary school and a good general report from the same school are prerequisites for admission to medical schools. In addition, applicants have to take written and oral tests in biology (10 papers), chemistry (10 papers), physics (10 papers), and socio-economic ideology. Further, following an interview with the chairman of the examination committee, the student's personality, cultural interests, and sociological attitudes and background are assessed. Double blind controls and daily random change of examiner groups are employed to minimize the effects of personal bias in the scoring of these tests and interviews. In the final stage of the selection process, all members of the committee are informed of the results for all applicants, and each member is entitled to award a bonus to any applicants he regards as especially promising. However, this bonus is limited to 2% of the candidate's score, and the total awarded to one candidate by all members must not exceed 10% of the score. The planned number of freshmen is then selected from those having scores above a certain level. This number includes about 10% who were initially rejected because of marginal grades but were eventually accepted after careful reappraisal. Mathematical and statistical analysis of the results thus far obtained and their correlation with the student's subsequent performance lead to the unequivocal conclusion that the written tests are superior to the oral examinations in the selection of successful candidates.

Assessment of student performance

At present, three types of assessment are used in Czechoslovakia.

1. "*Attestations*" *(course certificates)*. The junior teacher responsible for his "circle" (group) certifies that in the respective semester or year the student completed the prescribed course. Without these certificates the student cannot enrol for the following semester or year, nor apply for any examination.

2. *Examinations.* Each semester ends with examinations, and provision is made to schedule individual sessions at weekly intervals. Almost

all assessments of the student's knowledge are made by "orals", supplemented by practical tests wherever possible. The written "essay-type" examinations are limited to internal medicine, neurology and psychiatry and require preparation of a detailed record on patients examined by the student. The examinations in the various fields are usually developed by senior teachers, each of whom is responsible for his specialty.

Modern objective-type tests designed to evaluate knowledge, skills, understanding and problem-solving are at present under consideration but have not yet been introduced as obligatory. The good results, however, obtained with the multiple-choice questions in the admission tests have induced several medical schools in the country to adopt them in various specialties, particularly in basic sciences (e.g., in medical physics and histology).

3. *State examinations.* These examinations are conducted by a committee. Its chairman is nominated by the Ministry of Education and its members (professors) are appointed jointly by the rector of the university and the dean of the faculty. When appropriate, teachers of closely related disciplines are included.

In clinical disciplines, the examination is in two parts (practical and theorctical), and includes a complete examination of one or more patients and the interpretation of laboratory tests, X-ray pictures, clectrocardiograms, and ancillary data. In the oral examination the questions are written on cards, which are drawn at random by the student.

State examinations are required for internal medicine (during the 10th scmester), paediatrics (during the 10th semester), gynaecology and obstetrics (during the 11th or 12th semester), surgery (during the 11th or 12th semester), and public hygiene and social medicine (during the 11th and 12th semesters).

The grading system is uniform for all examinations; the grades are recorded as excellent, very good, good, or inadequate.

FRANCE

Relative uniformity

The organization of examinations in French medical faculties is regulated by laws, decrees and departmental orders of the National Ministry of Education. It might therefore be supposed that the systems of examination are nearly uniform in all the faculties. But this is not the case because the deans responsible for the general organization of the examination papers have the power to omit oral and practical tests or to modify them (essay-type or multiple-choice) to suit local conditions. Two examples follow:

(*a*) So-called "oral-written" tests, which were in fact written, essay-type tests, differing from the actual written test only in not preserving the student's anonymity, were instituted when the size of the classes in the Paris Medical School (1000-2000 students each year) made the practical organization of oral tests impossible.

(*b*) In 1962, after a trial period of two years, some faculties began to employ objective methods of examination such as multiple-choice questions. However, no special regulation has ever been issued by the National Ministry of Education about this, and officially the examination is still described as an essay examination. In fact, each department head is free to give either the traditional written examination (essay-type) or the objective (multiple-choice) type.

Thus, in spite of an apparent uniformity due to the control over teaching exercised by the National Ministry of Education, the diversity is considerable in the types of examination held in French medical schools.

Types of examination in current use

Various combinations of the following are in use:

(*a*) written examinations (traditional essay-type and/or multiple-choice)
(*b*) oral examinations
(*c*) practical tests
(*d*) examinations during internship *(stage clinique)*
(*e*) presentation of a thesis.

General conditions

Examinations are held yearly and there are two sessions (June and September). A student cannot take the same examination more than four times. In order to pass from the first into the second year and from the second into the third a student must reach a pass mark based on the average of the scores obtained in both basic science examinations and introduction to clinical subjects (*séméiologie clinique*). From the third year on, the student must reach a pass mark in each individual test in order to be promoted to the following year.

Examinations and competitive examinations (concours)

The reform of medical education (law of December 1958) has upheld the *externat* system as the basis of French hospital training,

but the selection procedures have changed. The non-resident students are no longer selected by means of special competitive examinations, but according to the level they have reached in the total examination grades of the first two years of the normal faculty curriculum. This system, which gives the examination grades a somewhat competitive value, calls for strict anonymity, identical questions for all the candidates, and a single board of examiners. This board is faced with a difficult task, for owing to the absence of a *numerus clausus* the number of students is very large and consequently each member of the board must read, evaluate and grade the same question (and there are 4-5 essay-type questions per student) a great many times.[1]

This situation has led some teachers to study objective methods of examination (e.g., multiple-choice tests) and to conclude that they had more advantages than disadvantages. As a result, multiple-choice examinations have been adopted in varying ways by many medical schools. At Lille, for example, they are used exclusively for all the yearly examinations of the first and second year (*externat*) as well as in the two subjects of the premedical year (CPEM); at Nancy, they are used by one department only (*physique médicale*); the situation in Paris is the same as in Lille except for the examinations in *séméiologie clinique;* at Montpellier, multiple-choice questions are used only in *sémiologie clinique* and for the second-year course of medical physics. Other schools are trying one or the other approach to objective tests as a means of evaluating students. This variation in the use of multiple-choice examinations indicates the extent of experimentation going on at present in all medical schools. It is to be hoped that those who are in favour of the new method of objective evaluation will have an opportunity to present arguments based on the advances made in educational psychology.

Future developments

The problem of examining large groups of medical students has stimulated interest in a movement that is likely to lead to the eventual establishment of departments of research in medical education in some medical schools. The fact that medical school faculties recognize the importance of examinations, not only as a means of evaluating students' knowledge but also as a guide for the teacher, supports this trend. A society of information and research in medical education was founded in October 1965 (SIREM—*La Société d'Information et de Recherche sur l'Education médicale*).

[1] In the provinces, medical faculties have to deal with 500 to 1000 first-year students per class. In Paris the number is approximately 2500, with 5000 students in the premedical year.

UNION OF SOVIET SOCIALIST REPUBLICS

Selection of students

Any student who possesses the *attestat zrjelosti* can apply for entry into any medical school within the USSR. In addition to the secondary school leaving certificate, the applicant must present a medical certificate and, if applicable, documents concerning the non-academic activities he has pursued after leaving the secondary school.[1] Selection among applicants is based on the results of an admission examination, which covers Russian literature, one foreign language, physics and chemistry; a good mark in the last two subjects is of greater value than a good mark in the first. This examination is held in every school throughout the country at the beginning of August. A commission, consisting of members of the medical faculty and local authorities, selects the requisite number of candidates according to their scores in this examination; in some cases, the quality of the secondary school leaving certificate is also considered. A student can apply only to one school; if he fails to gain admission to that school he can resubmit his application the following year to the same or another medical school.

Promotion and certification of students

At present there are two ways of controlling and evaluating the student's performance during the medical programme:

(*a*) by course certificates, which evaluate the work and knowledge shown by the students in practical courses;

(*b*) by examinations, which evaluate the total knowledge the students have acquired as a result of all forms of teaching.

Course certificates. In all disciplines that involve practical instruction (laboratory exercises, patient management, field or hospital experience, etc.), course certificates are required for admission into the next semester or year. The course certificate issued by the teacher responsible for the subject is added to the student's record (certificate book), usually without grading, since the certificate is based on continuous follow-up of the student during his practical work and is not, therefore, considered

[1] In addition, applicants who do not possess the secondary school leaving certificate but who have graduated as feldshers or nurses after having worked two or more years professionally can apply to sit for the admission examination. Once they have passed this examination, these applicants have priority in being admitted to the medical school as their previous experience is related to their future profession.

tantamount to an examination. In some schools or departments, minor examinations (written or oral) are included in the requirements for the course certificate and in such cases a full examination on the subject may be required of students whose regular performance has been inadequate. The so-called "classified course certificates" are no longer used since they were identical with examinations and therefore had the effect of unnecessarily increasing the total number of tests.

Examinations. The medical school sets the required level of knowledge in all disciplines in a manner designed to keep conditions for all students as nearly uniform as possible. Within this framework a list of questions is developed to cover all the major areas of the specialty; the number of questions varies with the nature and scope of the specialty. The questions are written on cards which are drawn at random by the students. These examination cards are designed to serve as an *aide-mémoire*, so that the examiner is neither obliged to rely exclusively on his memory nor to think up new questions on the spot. He is not rigidly bound by this system, however, and may ask other questions whenever he finds it necessary.

Written, oral or practical tests are generally given at various stages of the medical curriculum. The most frequently applied form is the "course examination" (*kursovyje ekzameny*), which is held immediately after instruction in a part of the subject matter has been completed. The students are examined by a single examiner, who need not be the senior staff member in charge of the course.

A more comprehensive form of evaluation is the "end-of-year examination" (*specialnyje perehodnyje ekzameny*), which consists of oral tests and laboratory exercises and, after the fourth year, of exercises in patient management. Successful performance entitles a student to pass from one academic year to another. This examination is held by the professor in charge of the theoretical course. If a student fails, he is given a chance to repeat the examination after one to three months.

State examinations (gosudarstvennyje)

1. *Preclinical examinations.* These are conducted by a committee, the chairman of which is nominated by the Ministry of Health of the Republic. Its members include teachers not only from the disciplines in which examinations are held but also from others, particularly the clinical sciences. Practical and theoretical examinations in anatomy and histology are taken after the third, and in physiology and biochemistry after the fourth semester. A student who is unsuccessful in one of these examinations is generally allowed only one chance to repeat it.

2. *Final examinations.* These are scheduled during the first two months of the twelfth semester and are confined to five disciplines: internal medicine; surgery; obstetrics and gynaecology; hygiene and organization of public health; and socio-economic sciences. In very broad disciplines, as for example internal medicine or surgery, an interim examination covering some limited aspects of the instruction may be offered. Each examination consists of both a practical and a theoretical part, the practical part being the more important.

The examinations are organized by a chairman who convenes a special committee composed of teachers in the disciplines concerned, teachers in other disciplines (including theoretical ones), and some medical practitioners. Students are graded as "excellent", "good", "adequate" and "inadequate". Those who fail may work as feldshers and/or repeat the examination a year later, at which time they are re-examined in all disciplines except those in which they were graded "excellent".

UNITED KINGDOM OF GREAT BRITAIN AND NORTHERN IRELAND

Selection of students

The choice of the university or medical school is left to the applicant, and multiple applications are common. In British universities a central clearing house has been established for allocating to other medical schools students who have failed to obtain their first choice. For the most part, qualification for enrolment in medical schools is based on secondary school examinations. However, when the number of admissions to a school is limited, various methods are used to screen applicants. In addition to entrance examinations (the minimum requirements of which may vary from school to school), competitive entrance scholarships in natural science, interviews, and reports from the secondary schools are often employed, although such reports have been shown to be poor indices of future performance.[1] Results of entrance examinations may also be used to award scholarships as, for example, in Glasgow, where a high correlation has been found between performance in these examinations and subsequent achievement.[2]

The strong emphasis on a science background as a prerequisite for entrance into a medical school suggests that opportunities are dwindling

[1] Furneaux, W. D. (1965) *The scientific background to university selection.* In: Reid, J. V. O. & Wilmot, A. J., ed., *Medical education in South Africa. Proceedings of the Conference on Medical Education, University of Natal, Durban, July 1964,* Pietermaritzburg, Natal University Press, pp. 197-211.
[2] Anderson, J. R., Lennox, B. & Low, A. (1964) *Lancet,* I, 96.

for the arts college student who wishes to become a physician. However, a student can transfer to medicine after obtaining an arts degree if he takes courses in chemistry, physics and mathematics, and several medical schools have arrangements whereby students can take such courses during the preclinical years. The alternative is for the student to obtain "A" level (advanced level) qualifications in these subjects in the general certificate of education examinations, at one of the many technical colleges that exist in the United Kingdom.

Assessment of student performance

In the premedical, preclinical and clinical years, a series of minor and major examinations, written, oral and practical, are used to determine the student's progress. Major examinations are held at the end of the first (premedical) year, the third (preclinical) year and the sixth (clinical) year. A year of compulsory work in hospital follows the final qualifying examination and, subject to approval by recognized teachers in medicine and surgery, the General Medical Council then awards registration. This registration licenses the doctor to practise. The doctor can go on to obtain an MD degree, by examination or by submitting a thesis about a research project.

Although some candidates are eliminated during the premedical year or in the final two years of "A" level studies at school, the rate of attrition during the first preclinical year (basic medical science) is still between 5% and 10%, depending on the university and school concerned.[1]

Also at the preclinical level, in some departments, such as anatomy, weekly oral tests are required. In other departments, such as physiology or biochemistry, records of performed experiments are accepted as sufficient evidence to allow the student to sit for the examination. In addition to these numerous examinations, some departments (at this level) require attendance at practical classes and lectures, while others adopt a more liberal attitude. Some universities place heavy emphasis on a tutorial system in which two or three students are assigned to each tutor, who is responsible for guiding the students and for giving them assistance when lectures or laboratories fail to provide adequate background or understanding. In these situations, students are encouraged to pursue their studies at an individual pace and the tutor is able to identify those who are likely to qualify for special honours at the end of their course. Other universities retain a system of course books which must be "signed up" before the student is allowed to progress to a further stage. Some students may be encouraged at this point to

[1] Perry, W. L. M. (1966) *Brit. J. med. Educ.*, 1, 16-24.

spend an extra year on a preclinical subject or in a general course, such as psychology or philosophy.

At the clinical level, too, the frequency of examinations varies from university to university. Very few are held in London by comparison with the large number required in Edinburgh. In addition to formal examinations, notes recorded by the chief of each relevant clinical section are used in most medical schools to document the student's progress through his various clinical assignments. Though generally of a subjective nature, these notes are sometimes based in part on the student's performance in written or practical examinations. The attrition rate in the clinical years is low: over 80% of the students pass the final examination at the first attempt; only 3–5% need to repeat it several times in order to attain qualification.

Licensure of physicians

Most universities and certain other bodies in the United Kingdom conduct qualifying examinations for doctors. The qualifying examinations, to which external examiners are usually called in by the respective faculty, have to meet with the approval of the General Medical Council (GMC). Other licensing bodies that do not fall under the jurisdiction of universities are the English Conjoint Board,[1] the Scottish Conjoint Board,[2] and the Society of Apothecaries of London, but these do not fall outside the jurisdiction of the GMC. In the London medical schools the Conjoint Board examinations are often taken concurrently with university qualifying examinations.

UNITED STATES OF AMERICA

Selection of students

Medical school applicants in the USA are not required to pass entrance examinations in course content; however, virtually all do take a Medical College Admissions Test (MCAT) designed to measure aptitude for dealing with verbal and quantitative concepts and achievement in science and humanities.

Studies of student performance have shown that the risk of failure is substantially higher among students with relatively low scores in the MCAT, particularly in the absence of some indication of compensatory strength. These same studies have also revealed a relatively low

[1] A joint examining body representing the Royal College of Physicians of London and the Royal College of Surgeons of England.

[2] A joint examining body representing the Royal Colleges of Physicians and Surgeons of Edinburgh and the Royal Faculty of Physicians and Surgeons of Glasgow.

correlation between MCAT scores and medical school grades, especially in the clinical years. However, each medical school is autonomous in determining its admission policy and the manner in which the results of this test are used by the admissions committees of different schools is quite variable: (*a*) they may, for all practical purposes, be completely disregarded; (*b*) a fairly rigid "cut-off" point may be employed, to be waived only in rare instances when a student appears to have other special qualifications; (*c*) the tests may be used as an aid in interpreting the prior academic records of students coming from many different premedical schools with divergent academic standards; or (*d*) in some schools the test may be used as an indicator of the need for more intensive investigation of students whose records reveal great discrepancies between scholastic aptitude and scholastic achievement.[1] Perhaps the one accurate generalization that can be made is that *no* school is inflexible to the point of depending on the MCAT as the sole determinant of admission.

Assessment of student performance

A recent survey[2] of examination practices in medical schools in the USA reveals great variation from school to school, and even from year to year. Within broad limits, departmental faculties bear sole responsibility, at both clinical and preclinical levels, for planning, constructing, scheduling, administering and grading examinations in the relevant disciplines. In most cases, therefore, separate disciplinary examinations are prepared for each subject. In a few schools these subject-oriented examinations are supplemented by a brief interdisciplinary section, which may or may not be given weight in assessing the student's overall progress. As of 1966, only one school reported a fully implemented system of comprehensive examinations prepared by a representative committee of the total faculty, a few others reported some partial development in this direction, and a few were planning to move toward such a system.

This detailed survey also disclosed that in most medical schools formal examinations are considered of great importance, especially at the preclinical level, in determining whether a student shall be permitted to continue the course. However, the nature and frequency of these formal examinations change over the course of the medical curriculum.

Nearly all schools indicated that, at the preclinical level, the formal examination is of far greater importance in the grading and promotion

[1] Tunkenstein, D. H. (1965) *J. med. Educ.*, 40, 1031.
[2] McGuire, C. (1966) *Survey of examination practices*, Center for the Study of Medical Education, University of Illinois, College of Medicine, Chicago, Illinois.

of students than are instructor's evaluations of the student's day-to-day work or of his habits and attitudes. At this level, the oral examination is used only occasionally, and external examiners are very rarely employed. Relatively few schools utilize essay-type examinations, and then only infrequently, during the preclinical programme, when the student's competence is most likely to be judged on the basis of the objective (multiple-choice) type of examination designed to assess his fund of knowledge. (About half the schools report a more or less exclusive reliance on this one type of examination at the preclinical level.) In some departments of some schools, these written examinations are supplemented by practical examinations of laboratory skills. In most schools, written and practical examinations are held very frequently during the preclinical years. Only four schools reported that each department administered a single such examination in *each* school term, and one school reported that it administered only one such examination in each discipline, at the end of the preclinical curriculum.

At the clinical level, the instructor's evaluations of the student's day-to-day performance in the clinic and of his professional habits and attitudes are likely to play a greater role than at the preclinical level, though most schools also consider the student's performance in some type of formal examination in determining his fate.

Oral examinations are much more common in the clinical than in the preclinical years, although even at this level external examiners are rarely asked to assist in administering them. Written examinations, too are important at the clinical level in most schools, and here again the essay type of examination is rarely used. In some schools the written and oral examinations are supplemented by a practical (bedside) examination, but clinical skills are most likely to be evaluated on the basis of the instructor's day-to-day contact with the student. Approximately half the schools report that the formal examinations at the clinical level are designed to determine with more or less equal emphasis "the student's fund of knowledge", his "capacity for solving realistic problems" and his "practical clinical skills". These examinations are commonly administered at the end of each school term, yet in some schools they may be held several times during each school term, while in others they are given only at the end of each clinical year or at the end of the entire clinical curriculum.

Licensure of physicians

Even though students in the USA receive the M.D. degree on completion of their formal medical school education, they are not licensed to practise medicine until they have successfully completed

one year of internship and have passed a state licensure examination. In most states a satisfactory grade in each part of the three-part examination prepared by the National Board of Medical Examiners is accepted as evidence that the requirements of the licensure examination have been met.

At present Part I of the National Board examination consists of a twelve-hour multiple-choice examination covering six preclinical disciplines (anatomy, bacteriology, biochemistry, pathology, pharmacology and physiology) and is normally taken at the end of the preclinical curriculum. Part II consists of a twelve-hour multiple-choice examination covering six major clinical specialties (medicine, surgery, paediatrics, obstetrics and gynaecology, psychiatry, and public health) and is normally taken at the end of the formal clinical curriculum. Parts I and II are administered to over 75% of all USA students *during* their medical school career. Part III, also an objective examination, is not administered until the end of the internship year; at present, it consists of two sections: (*a*) multiple-choice questions about the interpretation of films, X-ray photographs and other clinical data; (*b*) questions of an objective kind about the diagnostic investigation and management of selected cases. Formerly, a practical examination, designed to assess the candidate's ability to take a history, perform a physical examination and arrive at a diagnosis on two hospital patients, was included in the Part III examination. This has been discontinued because "the correlation between the ratings made by the two examiners who have evaluated this performance has been extremely low".[1]

---

[1] Schumacher, C. F. (1964) *J. med. Educ.*, 39, 192.

PUBLIC HEALTH PAPERS

No.		s. d.	$	Sw.fr.
1.	PSYCHIATRIC SERVICES AND ARCHITECTURE. *A. Baker, R. Llewelyn Davies & P. Sivadon* (1959) 59 pages	3/6	0.60	2.—
2.	EPIDEMIOLOGICAL METHODS IN THE STUDY OF MENTAL DISORDERS. *D. D. Reid* (1960) 79 pages. .	5/-	1.00	3.—
3.	HEALTH SERVICES IN THE USSR. Report Prepared by the Participants in a Study Tour Organized by the World Health Organization (1960) 58 pages	3/6	0.60	2.—
4.	ASPECTS OF PUBLIC HEALTH NURSING. *Various authors* (1961) 185 pages	8/6	1.75	5.—
5.	TRENDS IN JUVENILE DELINQUENCY. *T. C. N. Gibbens* (1961) 56 pages	3/6	0.60	2.—
6.	IONIZING RADIATION AND HEALTH. *Bo Lindell & R. Lowry Dobson* (1961) 81 pages	5/-	1.00	3.—
7.	BASIC NURSING EDUCATION PROGRAMMES. A GUIDE TO THEIR PLANNING. *Katharine Lyman* (1961) 81 pages	5/-	1.00	3.—
8.	THE ROLE OF IMMUNIZATION IN COMMUNICABLE DISEASE CONTROL. *Various authors* (1961) 118 pages	6/8	1.25	4.—
9.	TEACHING OF PSYCHIATRY AND MENTAL HEALTH. *Various authors* (1961) 186 pages	10/-	2.00	6.—
10.	CONTROL OF SOIL-TRANSMITTED HELMINTHS. *Paul C. Beaver* (1961) 44 pages	3/6	0,60	2.—
11.	MATERNAL AND CHILD HEALTH IN THE USSR. Report Prepared by the Participants in a Study Tour Organized by the World Health Organization (1962) 71 pages	5/-	1.00	3.—
12.	ROAD TRAFFIC ACCIDENTS. Epidemiology, Control and Prevention. *L. G. Norman* (1962) 110 pages. . . .	6/8	1.25	4.—
13.	ASPECTS OF WATER POLLUTION CONTROL. *Various authors* (1962) 165 pages	6/8	1.25	4.—
14.	DEPRIVATION OF MATERNAL CARE. A Reassessment of its Effects. *Various authors* (1962) 165 pages	10/-	2.00	6.—
15.	EPIDEMIOLOGY OF AIR POLLUTION. Report on a Symposium. *P. J. Lawther, A. E. Martin & E. T. Wilkins* (1962) 32 pages	1/9	0.30	1.—
16.	THE SCOPE OF EPIDEMIOLOGY IN PSYCHIATRY. *Tsung-Yi Lin & C. C. Stanley* (1962) 76 pages	5/-	1.00	3.—
17.	PAYING FOR HEALTH SERVICES. A Study of the Costs and Sources of Finance in Six Countries. *Brian Abel-Smith* (1963) 86 pages	5/-	1.00	3.—
18.	MEDICINE AND PUBLIC HEALTH IN THE ARCTIC AND ANTARCTIC. Selected Papers from a Conference. *Various authors* (1963) 169 pages	10/-	2.00	6.—

No.	s. d.	$	Sw.fr.
19. HEALTH EDUCATION IN THE USSR. Report Prepared by the Participants in a Study Tour Organized by the World Health Organization (1963) 69 pages	5/-	1.00	3.—
20. PREPARATION OF THE PHYSICIAN FOR GENERAL PRACTICE. *Various authors* (1963) 114 pages.	6/8	1.25	4.—
21. THE STAFFING OF PUBLIC HEALTH AND OUTPATIENT NURSING SERVICES. Methods of Study. *Doris E. Roberts* (1963) 100 pages	6/8	1.25	4.—
22. THE NURSE IN MENTAL HEALTH PRACTICE. Report on a Technical Conference. *Audrey L. John, Maria O. Leite-Ribeiro & Donald Buckle* (1963) 212 pages.	12/-	2.25	7.—
23. URBAN WATER SUPPLY CONDITIONS AND NEEDS IN SEVENTY-FIVE DEVELOPING COUNTRIES. *Bernd H. Dietrich & John M. Henderson* (1963) 92 pages	5/-	1.00	3.—
24. CARE OF CHILDREN IN DAY CENTRES. *Various authors* (1964) 189 pages	12/-	2.25	7.—
25. HOUSING PROGRAMMES: THE ROLE OF PUBLIC HEALTH AGENCIES. *Various authors* (1964) 197 pages	13/4	2.75	8.—
26. DOMESTIC ACCIDENTS. *E. Maurice Backett* (1965) 138 pages.	10/-	2.00	6.—
27. TRENDS IN THE STUDY OF MORBIDITY AND MORTALITY. *Various authors* (1965) 196 pages	13/4	2.75	8.—
28. ASPECTS OF FAMILY MENTAL HEALTH IN EUROPE. *Various authors* (1965) 123 pages	8/6	1.75	5.—
29. MASS CAMPAIGNS AND GENERAL HEALTH SERVICES. *C. L. Gonzalez* (1965) 87 pages.	6/8	1.25	4.—
30. NOISE. An Occupational Hazard and Public Nuisance. *Alan Bell* (1966) 131 pages.	10/-	2.00	6.—
31. A GUIDE FOR STAFFING A HOSPITAL NURSING SERVICE. *Marguerite Paetznick* (1966) 93 pages. . .	6/8	1.25	4.—
32. AN INTERNATIONAL STUDY OF HEALTH EXPENDITURE AND ITS RELEVANCE FOR HEALTH PLANNING. *Brian Abel-Smith* (1967) 127 pages	12/-	2.00	6.—
33. THE PHYSIOLOGICAL BASIS OF HEALTH STANDARDS FOR DWELLINGS. *M. S. Goromosov* (1968) 99 pages	10/-	1.75	5.—
34. PRINCIPLES AND PRACTICE OF SCREENING FOR DISEASE. *J. M. G. Wilson & G. Jungner* (1968) 163 pages	14/-	2.25	7.—
35. PREVENTION OF SUICIDE (1968) 84 pages	8/-	1.25	4.—
36. A REVIEW OF THE NATURE AND USES OF EXAMINATIONS IN MEDICAL EDUCATION. *J. Charvat, C. McGuire & V. Parsons* (1968) 74 pages	10/-	1.75	5.—

WHO publications may be obtained through:

AFGHANISTAN	*see* India, WHO Regional Office.
ARGENTINA	Editorial Sudamericana S.A., Humberto 1° 545, BUENOS AIRES.
AUSTRALIA	Hunter Publications, 23 McKillop Street, MELBOURNE C. 1 — United Nations Association of Australia, Victorian Division, 364 Lonsdale Street, MELBOURNE, Victoria 3000.
AUSTRIA	Gerold & Co., I. Graben 31, VIENNA 1.
BELGIUM	Office international de Librairie, 30 av. Marnix, BRUSSELS.
BURMA	*see* India, WHO Regional Office.
CAMBODIA	The WHO Representative, P.O. Box 111, PHNOM-PENH.
CANADA	The Queen's Printer, OTTAWA.
CEYLON	*see* India, WHO Regional Office.
CHINA	The WHO Representative, 5 Chungshan Road South, TAIPEI, TAIWAN — The World Book Co., Ltd, 99 Chungking South Road, Section 1, TAIPEI, Taiwan.
COLOMBIA	Distrilibros Ltd, Pio Alfonso García, Carrera 4a, Nos 36-119, CARTAGENA.
CONGO	Librairie congolaise, 12 avenue des Aviateurs, KINSHASA.
COSTA RICA	Imprenta y Librería Trejos S.A., Apartado 1313, SAN JOSÉ.
CYPRUS	MAM, P.O. Box 1674, NICOSIA.
DENMARK	Ejnar Munksgaard, Ltd, Nørregade 6, COPENHAGEN.
ECUADOR	Librería Científica S.A., P.O. Box 362, Luque 223, GUAYAQUIL.
FIJI	The WHO Representative, P.O. Box 113, SUVA.
FINLAND	Akateeminen Kirjakauppa, Keskuskatu 2, HELSINKI.
FRANCE	Librairie Arnette, 2 rue Casimir-Delavigne, PARIS 6e.
GERMANY	Govi-Verlag GmbH, Beethovenplatz 1-3, FRANKFURT A. M. 6 — W. E. Saarbach, Postfach 1510, Follerstrasse 2, COLOGNE 1 — Alex Horn, Spiegelgasse 9, WIESBADEN.
GREECE	G. C. Eleftheroudakis S.A., Librairie internationale, rue Nikis 4, ATHENS (T. 126).
HAITI	Max Bouchereau, Librairie "A la Caravelle", Boîte postale 111-B, PORT-AU-PRINCE.
HUNGARY	Kultura, P.O.B. 149, BUDAPEST 62 — Akadémiai Könyvesbolt, Váci utca 22, BUDAPEST V.
ICELAND	Snaebjörn Jonsson & Co., P.O. Box 1131, Hafnarstraeti 9, REYKJAVIK.
INDIA	WHO Regional Office for South-East Asia, World Health House, Indraprastha Estate, Ring Road, NEW DELHI 1 — Oxford Book & Stationery Co., Scindia House, NEW DELHI; 17 Park Street, CALCUTTA 16 (Sub-agent).
INDONESIA	*see* India, WHO Regional Office.
IRAN	Mebso Bookstore, Naderi Avenue (Arbab-Guiv Building), TEHERAN.
IRELAND	The Stationery Office, DUBLIN.
ISRAEL	Heiliger & Co., 3 Nathan Strauss Street, JERUSALEM.
ITALY	Edizioni Minerva Medica, Corso Bramante 83-85, TURIN; Via Lamarmora 3, MILAN.
JAPAN	Maruzen Company, Ltd, 6 Tori-Nichome Nihonbashi, TOKYO.
KENYA	The Caxton Press Ltd, Head Office: Gathani House, Huddersfield Road, P.O. Box 1742, NAIROBI.
KOREA	The WHO Representative, Central P.O. Box 540, SEOUL.
LAOS	The WHO Representative, P.O. Box 343, VIENTIANE.
LEBANON	Librairie Au Papyrus, Immeuble Abdel Baki, rue Cinéma Colisée, Hamra, BEIRUT.
LUXEMBOURG	Librairie Trausch-Schummer, place du Théâtre, LUXEMBOURG.
MALAYSIA	The WHO Representative, P.O. Box 2550, KUALA LUMPUR — Jubilee (Book) Store Ltd, 97 Batu Road, KUALA LUMPUR.
MEXICO	La Prensa Médica Mexicana, Ediciones Científicas, Paseo de las Facultades 26, MEXICO CITY 20, D.F.
MONGOLIA	*see* India, WHO Regional Office.
MOROCCO	Editions La Porte, 281 avenue Mohammed V, RABAT.

WHO publications may be obtained through:

NEPAL	*see* India, WHO Regional Office.
NETHERLANDS	N.V. Martinus Nijhoff's Boekhandel en Uitgevers Maatschappij, Lange Voorhout 9, THE HAGUE.
NEW ZEALAND	Government Printing Office, Government Bookshops at State Advances Building, Rutland Street (P.O. Box 5344), AUCKLAND; 20 Molesworth Street (Private Bag), WELLINGTON; 112 Gloucester Street (P.O. Box 1721), CHRISTCHURCH; Stock Exchange Building, Princes Street (P.O. Box 1104), DUNEDIN — R. Hill & Son Ltd, Ideal House, Cnr. Gilles Avenue & Eden St., Newmarket, AUCKLAND S.E. 1.
NIGERIA	University Bookshop Nigeria Ltd, University of Ibadan, IBADAN.
NORWAY	Johan Grundt Tanum Forlag, Karl Johansgt. 41, OSLO.
PAKISTAN	Mirza Book Agency, 65 Shahrah Quaid-E. Azam, P.O. Box 729, LAHORE 3 — Shilpa Niketan, 29 D.I.T. Super Market, Mymensingh Road, DACCA 2.
PARAGUAY	Agencia de Librerias Nizza S.A., Estrella No. 721, ASUNCIÓN.
PERU	Distribuidora Inca S.A., Apartado 3115, Emilio Althaus 470, LIMA.
PHILIPPINES	World Health Organization, Regional Office for the Western Pacific, P.O. Box 2932, MANILA.
POLAND	Skladnica Ksiegarska ul. Mazowiecka 9, WARSAW *(except periodicals)* — BKWZ Ruch, ul. Wronia 23, WARSAW *(periodicals only)*.
PORTUGAL	Livraria Rodrigues, 186 Rue Aurea, LISBON.
SINGAPORE	City Book Store Ltd, Winchester House, Collyer Quay.
SOUTH AFRICA	Van Schaik's Bookstore (Pty) Ltd, P.O. Box 724, PRETORIA.
SPAIN	Comercial Atheneum S.A., Consejo de Ciento 130-136, BARCELONA 15; Vergara 9, MADRID — Librería Díaz de Santos, Lagasca 38, MADRID 1.
SWEDEN	Aktiebolaget C.E. Fritzes Kungl. Hovbokhandel, Fredsgatan 2, STOCKHOLM 16.
SWITZERLAND	Medizinischer Verlag Hans Huber, Marktgasse 9, BERNE.
THAILAND	*see* India, WHO Regional Office.
TUNISIA	Société Tunisienne de Diffusion, 5 avenue de Carthage, TUNIS.
TURKEY	Librairie Hachette, 469 av. de l'Indépendance, ISTANBUL.
UGANDA	*see address under* KENYA.
UNITED KINGDOM	H.M. Stationery Office: 49 High Holborn, LONDON W.C. 1; 13a Castle Street, EDINBURGH 2; 109 St Mary Street, CARDIFF CF1, 1JW; 7-11 Linenhall Street, BELFAST BT2, 8AY; Brazennose Street, MANCHESTER 2; 258-259 Broad Street, BIRMINGHAM 1; 50 Fairfax Street, BRISTOL 1. *All postal orders should be sent to P.O. Box 569, London S.E.1.*
UNITED REPUBLIC OF TANZANIA	*see address under* KENYA.
UNITED STATES OF AMERICA	The American Public Health Association, Inc., 1740 Broadway, NEW YORK, N.Y. 10019.
USSR	*For readers in the USSR requiring Russian editions:* Komsomolskij prospekt 18, Medicinskaja Kniga, MOSCOW — *For readers outside the USSR requiring Russian editions:* Kuzneckij most 18, Mezdunarodnaja Kniga, MOSCOW G-200.
VENEZUELA	The University Society Venezolana C.A., Apartado 10786, CARACAS — Librería del Este, Av. Francisco de Miranda 52, Edificio Galipán, CARACAS.
VIET-NAM	The WHO Representative, P.O. Box 242, SAIGON.
YUGOSLAVIA	Državno Preduzeče Jugoslovenska Knjiga, Terazije 27/II, BELGRADE.

Orders may also be addressed to: World Health Organization, Distribution and Sales Unit, Geneva, Switzerland, but must be paid for in pounds sterling, US dollars, or Swiss francs.

Price: 10/- $1.75 Sw. fr. 5.—